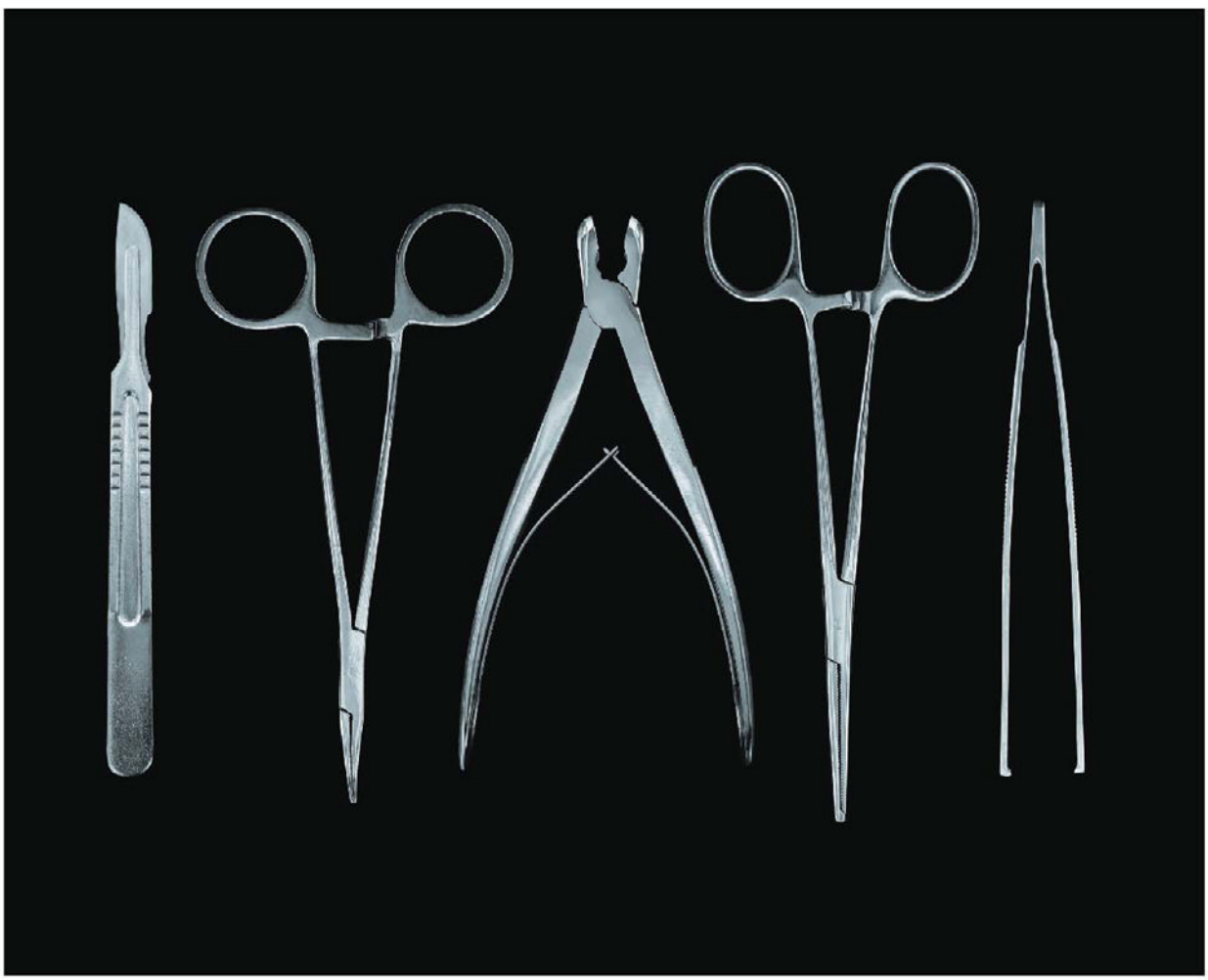
ROUTLEDGE

The Routledge Companion to Philosophy of Language



Edited by Gillian Russell
and Delia Graff Fara

THE ROUTLEDGE COMPANION TO PHILOSOPHY OF LANGUAGE

Philosophy of language is the branch of philosophy that examines the nature of meaning, the relationship of language to reality, and the ways in which we use, learn, and understand language.

The Routledge Companion to Philosophy of Language provides a comprehensive and up-to-date survey of the field, charting its key ideas and movements, and addressing contemporary research and enduring questions in the philosophy of language. Unique to this *Companion* is clear coverage of research from the related disciplines of formal logic and linguistics, and discussion of the applications in metaphysics, epistemology, ethics, and philosophy of mind.

Organized thematically, the *Companion* is divided into seven sections: Core Topics; Foundations of Semantics; Parts of Speech; Methodology; Logic for Philosophers of Language; Philosophy of Language for the Rest of Philosophy; and Historical Perspectives.

Comprised of 70 essays from leading scholars—including Sally Haslanger, Jeffrey King, Sally McConnell-Ginet, Rae Langton, Kit Fine, John MacFarlane, Jeff Pelletier, Scott Soames, Jason Stanley, Stephen Stich and Zoltán Gendler Szabó—*The Routledge Companion to Philosophy of Language* promises to be the most comprehensive and authoritative resource for students and scholars alike.

Gillian Russell is Associate Professor of Philosophy at Washington University in St. Louis.

Delia Graff Fara is Associate Professor of Philosophy at Princeton University.

Routledge Philosophy Companions

Routledge Philosophy Companions offer thorough, high quality surveys and assessments of the major topics and periods in philosophy. Covering key problems, themes and thinkers, all entries are specially commissioned for each volume and written by leading scholars in the field. Clear, accessible and carefully edited and organised, *Routledge Philosophy Companions* are indispensable for anyone coming to a major topic or period in philosophy, as well as for the more advanced reader.

The Routledge Companion to Aesthetics, Second Edition
Edited by Berys Gaut and Dominic Lopes

The Routledge Companion to Philosophy of Religion
Edited by Chad Meister and Paul Copan

The Routledge Companion to the Philosophy of Science
Edited by Stathis Psillos and Martin Curd

The Routledge Companion to Twentieth Century Philosophy
Edited by Dermot Moran

The Routledge Companion to Philosophy and Film
Edited by Paisley Livingston and Carl Plantinga

The Routledge Companion to Philosophy of Psychology
Edited by John Symons and Paco Calvo

The Routledge Companion to Metaphysics
Edited by Robin Le Poidevin, Peter Simons, Andrew McGonigal, and Ross Cameron

The Routledge Companion to Nineteenth Century Philosophy
Edited by Dean Moyar

The Routledge Companion to Ethics
Edited by John Skorupski

The Routledge Companion to Epistemology
Edited by Sven Bernecker and Duncan Pritchard

The Routledge Companion to Philosophy and Music
Edited by Theodore Gracyk and Andrew Kania

The Routledge Companion to Phenomenology
Edited by Søren Overgaard and Sebastian Luft

The Routledge Companion to Philosophy of Language
Edited by Gillian Russell and Delia Graff Fara

Forthcoming:
The Routledge Companion to Seventeenth Century Philosophy
Edited by Dan Kaufman

The Routledge Companion to Eighteenth Century Philosophy
Edited by Aaron Garrett

The Routledge Companion to Philosophy of Mental Disorder
Edited by Jakob Hohwy and Philip Gerrans

The Routledge Companion to Social and Political Philosophy
Edited by Gerald Gaus and Fred D'Agostino

The Routledge Companion to Theism
Edited by Charles Taliaferro, Victoria Harrison, and Stewart Goetz

The Routledge Companion to Philosophy of Law
Edited by Andrei Marmor

The Routledge Companion to Islamic Philosophy
Edited by Richard C. Taylor and Luis Xavier López-Farjeat

The Routledge Companion to Philosophy of Literature
Edited by Noël Carroll and John Gibson

The Routledge Companion to Philosophy of Religion, Second Edition
Edited by Chad Meister and Paul Copan

The Routledge Companion to Aesthetics, Third Edition
Edited by Berys Gaut and Dominic Lopes

The Routledge Companion to Philosophy of Science, Second Edition
Edited by Stathis Psillos and Martin Curd

The Routledge Companion to Ancient Philosophy
Edited by Frisbee Sheffield and James Warren

PRAISE FOR THE SERIES

The Routledge Companion to Aesthetics

'This is an immensely useful book that belongs in every college library and on the bookshelves of all serious students of aesthetics.'—***Journal of Aesthetics and Art Criticism***

'The succinctness and clarity of the essays will make this a source that individuals not familiar with aesthetics will find extremely helpful.'—***The Philosophical Quarterly***

'An outstanding resource in aesthetics . . . this text will not only serve as a handy reference source for students and faculty alike, but it could also be used as a text for a course in the philosophy of art.—***Australasian Journal of Philosophy***

'Attests to the richness of modern aesthetics . . . the essays in central topics – many of which are written by well-known figures – succeed in being informative, balanced and intelligent without being too difficult.'—***British Journal of Aesthetics***

'This handsome reference volume . . .belongs in every library.'—***Choice***

'The *Routledge Companion*s to Philosophy have proved to be a useful series of high quality surveys of major philosophical topics and this volume is worthy enough to sit with the others on a reference library shelf.'—***Philosophy and Religion***

The Routledge Companion to Philosophy of Religion

'. . . a very valuable resource for libraries and serious scholars.'—***Choice***

'The work is sure to be an academic standard for years to come . . . I shall heartily recommend *The Routledge Companion to Philosophy of Religion* to my students and colleagues and hope that libraries around the country add it to their collections.'—***Philosophia Christi***

The Routledge Companion to Philosophy of Science

A ***Choice*** Outstanding Academic Title 2008
'With a distinguished list of internationally renowned contributors, an excellent choice of topics in the field, and well-written, well-edited essays throughout, this compendium is an excellent resource. Highly recommended.'—***Choice***

'Highly recommended for history of science and philosophy collections.'—***Library Journal***

This well conceived companion, which brings together an impressive collection of distinguished authors, will be invaluable to novices and experience readers alike.' —***Metascience***

The Routledge Companion to Twentieth Century Philosophy

'To describe this volume as ambitious would be a serious understatement. . . . full of scholarly rigor, including detailed notes and bibliographies of interest to professional philosophers. . . . Summing up: Essential.'—***Choice***

The Routledge Companion to Philosophy and Film

'A fascinating, rich volume offering dazzling insights and incisive commentary on every page . . . Every serious student of film will want this book . . . Summing Up: Highly recommended.'—***Choice***

The Routledge Companion to Philosophy of Psychology

'This work should serve as the standard reference for those interested in gaining a reliable overview of the burgeoning field of philosophical psychology. Summing Up: Essential.'—***Choice***

The Routledge Companion to Metaphysics

'The *Routledge Philosophy Companions* series has a deserved reputation for impressive scope and scholarly value. This volume is no exception . . . Summing Up: Highly recommended.'—***Choice***

The Routledge Companion to Nineteenth Century Philosophy

A ***Choice*** Outstanding Academic Title 2010
'This is a crucial resource for advanced undergraduates and faculty of any discipline who are interested in the 19th-century roots of contemporary philosophical problems. Summing Up: Essential.'—***Choice***

The Routledge Companion to Ethics

'This fine collection merits a place in every university, college, and high school library for its invaluable articles covering a very broad range of topics in ethics[.] . . . With its remarkable clarity of writing and its very highly qualified contributors, this volume is must reading for anyone interested in the latest developments in these important areas of thought and practice. Summing Up: Highly recommended.'—***Choice***

THE ROUTLEDGE COMPANION TO PHILOSOPHY OF LANGUAGE

Edited by
Gillian Russell and Delia Graff Fara

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2012
by Routledge
711 Third Avenue, New York, NY 10017

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
The Routledge companion to philosophy of language/edited by Delia Graff Fara and Gillian Russell.
p. cm.—(Routledge Philosophy Companions)
Includes bibliographical references and index.
1. Language and languages—Philosophy. I. Fara, Delia Graff. II. Russell, Gillian.
P106.R75 2011
121'.68—dc23
2011023340

ISBN: 978–0–415–99310–4 (hbk)
ISBN: 978–0–203–20696–6 (ebk)

Typeset in Goudy by Swales & Willis Ltd, Exeter, Devon

CONTENTS

NOTES ON CONTRIBUTORS

Barbara Abbott is a Michigan State University Professor Emerita of Linguistics and Philosophy of Language. She has worked on such topics as definiteness, proper names, presuppositions and conditionals, and is the author of *Reference* (Oxford University Press, 2010).

Luvell Anderson is the Alain Locke Postdoctoral Fellow in the Department of Philosophy at Rutgers University. He works on issues in the philosophy of language and race.

Aldo Antonelli is professor of philosophy at the University of California, Davis. His research interests include topics in the philosophy of logic and mathematics, defeasible logics and knowledge representation, philosophical logic, and the theory of generalized quantifiers.

Kent Bach, Professor Emeritus of Philosophy at San Francisco State University, was educated at Harvard College and University of California, Berkeley. He has written extensively in philosophy of language, theory of knowledge, and philosophy of mind. His books include *Thought and Reference* (Oxford University Press, 1987, expanded edition 1994) and, with Michael Harnish, *Linguistic Communication* and Speech Acts (MIT Press, 1979).

Michael Beaney is Professor of Philosophy at the University of York, United Kingdom. He works on the history of analytic philosophy and on conceptions of analysis in the history of philosophy. He is the author of *Frege: Making Sense* (Duckworth, 1996), and editor of *The Frege Reader* (Blackwell, 1997), *Gottlob Frege: Critical Assessments of Leading Philosophers* (with Erich Reck; 4 vols., Routledge, 2005), *The Analytic Turn* (Routledge, 2007), and *The Oxford Handbook of the History of Analytic Philosophy* (Oxford University Press, forthcoming).

Johan van Benthem is University Professor of Logic at the University of Amsterdam, and Professor of Philosophy at Stanford University. His current interests are dynamic logics of information, agency, and games. Relevant works include *Exploring Logical Dynamics* (CSLI Publications, 1996) and *Logical Dynamics of Information and Interaction* (Cambridge University Press 2010), and he was a co-editor of the *Handbook of Logic and Language* (Elsevier Amsterdam, 1997, 2010).

Alexander Bird is Professor of Philosophy at the University of Bristol. His research concerns the metaphysics of science, in particular dispositions, natural kinds, and laws of nature. He is the author of *Nature's Metaphysics* (Oxford University Press, 2007) and is now working on two projects, one concerning natural kinds and and the other concerning scientific knowledge.

David Braun is a Professor of Philosophy at University at Buffalo. His main research interests are in philosophy of language and philosophy of mind.

John P. Burgess has taught since 1975 at Princeton University, where he is now John N. Woodhull Professor of Philosophy. He has worked mainly in logic and allied areas, and is author or co-author of *A Subject with No Object* (with Gideon Rosen); *Computability and Logic* (with George Boolos and Richard Jeffrey); *Fixing Frege*; *Mathematics, Models, and Modality*; *Philosophical Logic*; and *Truth* (with Alexis G. Burgess).

Daniel Büring is Professor of Linguistics at the University of Vienna, Austria. His main research areas are semantics, pragmatics, and intonation; his publications include *The Meaning of Topic and Focus* (1997, Routledge), *Binding Theory* (Cambridge University Press, 2005) and the forthcoming *Intonation and Meaning*.

Robyn Carston is Professor of Linguistics at University College London. She works on the semantics/pragmatics distinction, the explicit/implicit communication distinction, and the interpretation of metaphor. She has published *Thoughts and Utterances* (Blackwell, 2002) and a collection of her papers *Pragmatics and Semantic Content* is forthcoming (Oxford University Press).

Luca Castagnoli is Lecturer in Ancient Philosophy at the Department of Classics and Ancient History Durham University. His research interests include ancient logic and epistemology and he is the author of *Ancient Self-Refutation* (Cambridge University Press, 2010).

Paul Dekker is Associate Professor at the ILLC/Department of Philosophy, of the Universiteit van Amsterdam. His work originates in the philosophy of language and formal semantics. More specific interests include the dynamic semantics, the semantics/pragmatic interface, quantification and modality. He has edited a book on the semantics and pragmatics of questions, and is currently completing a monograph on an original interpretation of formal dynamic semantics.

Josh Dever is Associate Professor of Philosophy at the University of Texas at Austin. He works in the philosophy of language, philosophical logic, and metaphysics.

Michael Devitt is Distinguished Professor of Philosophy at the Graduate Center at CUNY. His research interests include the philosophy of linguistics and the philosophy of language. He is the author of *Designation* (Columbia, 1981), *Coming to Our Senses* (Cambridge University Press, 1996), *Language and Reality* (with Kim Sterelny, MIT Press, 1999), *Ignorance of Language* (Oxford University Press 2006), and the editor (with Richard Hanley) of the *Blackwell Guide to the Philosophy of Language* (Blackwell, 2006). His most recent book is *Putting Metaphysics First* (Oxford University, Press, 2010).

Sinan Dogramaci is Assistant Professor of Philosophy at the University of Texas at Austin. Most of his recent research concerns questions about how we can rationally infer conclusions that are logical consequences of things we already know.

Tom Donaldson is a PhD student at Rutgers University. He is interested in meaning and reasoning, and is working on a dissertation on pragmatism.

Matti Eklund is Associate Professor of Philosophy at Cornell University. His published work is primarily on topics in metaphysics, philosophy of language, and philosophy of logic.

Kit Fine is University Professor and Silver Professor of Philosophy and Mathematics at New York University. His principal interests are metaphysics, logic, and the philosophy of language.

Kai von Fintel is Professor of Linguistics in the Department of Linguistics and Philosophy at the Massachusetts Institute of Technology and Associate Dean of MIT's School of Humanities, Arts, and Social Sciences. He is editor of the journal *Semantics & Pragmatics*. His research is primarily in semantics, pragmatics, philosophy of language, and the intersections thereof. He has worked on modality, conditionals, presuppositions, quantification, context-dependency, and negative polarity.

Robert Freidin is Professor of Linguistics in the Council of the Humanities at Princeton University. His research concerns syntax and semantics, focusing on the foundations of syntactic theory (the central concepts of syntactic analysis and their evolution) and their role in the study of language and mind. Some of this work is collected in *Generative Grammar: Theory and its History* (Routledge, 2007). His most recent publications include "The roots of minimalism" (with Howard Lasnik) in *The Oxford Handbook of Linguistic Minimalism*, "Noam Chomsky's contribution to linguistics: a preliminary sketch" in *The Oxford Handbook of the History of Linguistics and Syntax: basic concepts and applications* (forthcoming from Cambridge University Press).

Antony Galton is Reader in Knowledge Representation in the College of Engineering, Mathematics, and Physical Sciences at the University of Exeter. He has published widely on the topic of spatial, temporal, and spatio-temporal knowledge representation, with particular emphasis on the conceptual foundations of geographical information science. Earlier work explored the logical and philosophical underpinning of aspectual phenomena in natural language.

Christopher Gauker is Professor of Philosophy at the University of Cincinnati. He is the author of *Thinking Out Loud: An Essay on the Relation between Thought and Language* (Princeton, 1994), *Words without Meaning* (MIT Press, 2003), *Conditionals in Context* (MIT Press 2005) and *Words and Images: An Essay on the Origin of Ideas* (Oxford University Press, 2011).

Stavroula Glezakos is Associate Professor of Philosophy at Wake Forest University. Her research focuses on questions concerning the meaning and reference of names and kind terms.

Delia Graff Fara is an Associate Professor in the Philosophy Department at Princeton University. Her research and teaching interests fall mainly in the areas of philosophy of language, metaphysics, philosophical logic, and philosophical linguistics.

Anthony S. Gillies is an Associate Professor of Philosophy at Rutgers University. His research focuses on issues about language, logic, and information. He has worked on conditionals, modality, and the dynamics of both meaning and belief.

Paul Hagstrom is Associate Professor of Linguistics in the Department of Romance Studies at Boston University. His research centers around the syntax, morphology, and semantics of interrogative sentences and indefinites, and he also works on first language acquisition of morphosyntax and second language acquisition of definiteness restrictions. He has published overviews of interrogative semantics (2006, Glot International, 2003) and of A-not-A questions (*Blackwell Companion to Syntax*, 2006).

Sally Haslanger is Professor of Philosophy and Director of Women's and Gender Studies at MIT. She specializes in feminist theory and critical race theory with an emphasis on questions concerning metaphysics, epistemology, and philosophy of language.

Allen Hazen is a Fellow of the University of Melbourne's philosophy department and Professor of Philosophy at the University of Alberta. His research interests include metaphysics, philosophy of mathmatics, logic, philosophy of language, philosophy of mind, epistemology, and history of philosophy.

Laurence Horn is Professor of Linguistics and Philosophy at Yale University. He is the author of *A Natural History of Negation* and numerous papers on negation and pragmatics and the (co-)editor of *Negation and Polarity* (Oxford University Press, 2000), *The Handbook of Pragmatics* (Blackwell), *Explorations in Pragmatics* (de Gruyter, 2001) and *The Expression of Negation* (de Gruyter, 2010).

Cory Juhl is Professor of Philosophy at the University of Texas at Austin. His research areas include the philosophy of mathematics, metaphysics, and metasemantics. His recent papers include works on the fine-tuning argument, on the inverse gambler's fallacy, and another paper on the distinction between theoretical entities and 'stipulata'. He is co-author of *Analyticity* (2010, Routledge).

Chris Kennedy is Professor and Chair of the Department of Linguistics at the University of Chicago. His research addresses questions in semantics, pragmatics, and philosophy of language, primarily through an exploration of the grammar of comparison, amount, degree, and vagueness. He is the general editor (with Chris Barker) of Oxford University Press's *Studies in Semantics and Pragmatics* and *Surveys in Semantics and Pragmatics*.

Jeffrey C. King Professor in the Philosophy Department at Rutgers University. His research focuses on the philosophy of language, formal semantics, philosophical logic and metaphysics. He is the author of two books: *The Nature and Structure of Content* (Oxford University Press, 2007) and *Complex Demonstratives: A Quantificational Account*. (MIT Press, 2001).

Gyula Klima is Professor of Philosophy at Fordham University, New York, and Director of the Society for Medieval Logic and Metaphysics. His books include *John Buridan* (Oxford University Press, 2009), *Readings in Medieval Philosophy* (Blackwell, 2007), *John Buridan: Summulae de Dialectica*, an annotated translation with a philosophical introduction (Yale University Press, 2001), and *ARS ARTIUM: Essays in Philosophical Semantics, Medieval and Modern* (Hungarian Academy of Sciences, 1988).

Stephen T. Kuhn is a professor in the Department of Philosophy and an affiliated faculty member in the Department of Linguistics at Georgetown University. His interests include modal and tense logic, philosophy of logic, and the evolution of moral norms.

Rae Langton is Professor of Philosophy at MIT. She works on moral and political philosophy, feminist philosophy, speech act theory, metaphysics, and the history of philosophy.

Ermelinda Valentina Di Lascio is Leverhulme Early Career Fellow at the Department of Classics and Ancient History of Durham University. Her research interests include ancient logic and dialectic and she has published on Aristotle's Sophistical Refutations, logical papyri, and Wittgenstein.

Ernie Lepore is a professor of philosophy at Rutgers University. He works in philosophy of language.

Sarah-Jane Leslie is an Assistant Professor in the Philosophy Department at Princeton University. She is also associated faculty in the Psychology Department, and serves on the executive committee for the Program in Linguistics. Her work is primarily in the intersection between philosophy and psychology. She is currently writing a book on generics for Oxford University Press, which will bring issues in philosophy of language into contact with recent empirical results from psychology.

Eric Loomis is Associate Professor of Philosophy at the University of South Alabama. His research areas include metaphysics and early analytic philosophy. He is co-author, with Cory Juhl, of *Analyticity* (Routledge, 2010).

Michael Losonsky is Professor of Philosophy at Colorado State University. He is author of *Linguistic Turns in Modern Philosophy* (Cambridge University Press, 2006), *Enlightenment and Action from Descartes to Kant* (Cambridge University Press, 2001) and co-editor with H. Geirsson of *Readings in Language and Mind* (Blackwell, 1996). He works on the philosophies of Locke and Leibniz, the history of the philosophy of language, and the history of semantic internalism.

Peter Ludlow (PhD Columbia U, 1985) taught at the state University of New York at Stony Brook, the University of Michigan, and the University of Toronto before joining Northwestern University. He is the editor of the *Readings in the Philosophy of Language* (Bradford Books, 1997). Lately he has been working on a number of topics including the semantics of quantified noun-phrases, microlanguages and the dynamic

lexicon and natural logic. His most recent book is *The Philosophy of Generative Linguistics* (Oxford University Press, 2011).

Kirk Ludwig is Professor of Philosophy at Indiana University, Bloomington. He publishes in the philosophy of language, mind, action (esp. collective action), metaphysics and epistemology. He is editor of *Donald Davidson* (Cambridge University Press, 2003) and co-author with Ernie Lepore of *Donald Davidson: Meaning, Truth, Language and Reality* (Oxford University Press, 2005) and *Donald Davidson's Truth-theoretic Semantics* (Oxford University Press, 2007).

John MacFarlane is Professor of Philosophy at the University of California, Berkeley. He has written extensively on the philosophy of logic and the philosophy of language, and is currently working on a book on relative truth and its applications.

Edouard Machery is Associate Professor of History and Philosophy of Science in the Department of History and Philosophy of Science at the University of Pittsburgh, a resident fellow of the Center for Philosophy of Science (University of Pittsburgh), a member of the Center for the Neural Basis of Cognition (Carnegie Mellon University and University of Pittsburgh), and an associate editor of the *European Journal for Philosophy of Science*. He is the author of *Doing without Concepts* (Oxford University Press, 2009).

Robert May is Professor of Philosophy and Linguistics at the University of California, Davis. He is well-known for his research in theoretical linguistics, especially on the syntax and semantics of quantification, and is currently working in philosophy of language and linguistics, philosophy of logic, and history of early analytic philosophy. His books include *Logical Form: Its Structure and Derivation* (MIT Press, 1985), *Indices and Identity* (MIT Press, 1994) and *De Lingua Belief* (MIT Press, 2006), the last two co-authored with Robert Fiengo.

Sally McConnell-Ginet, who writes on various topics in semantics, pragmatics, and the philosophy of language, is Professor Emerita of Linguistics, Cornell. *Gender, Sexuality, and Meaning* (Oxford University Press, 2011) and her other work on the interaction of language with gender and sexuality draw on linguistic and philosophical perspectives.

Michael McDermott studied philosophy at the University of Sydney under David Armstrong, and has taught there ever since.

Lawrence S. Moss is Director of the Program in Pure and Applied Logic at Indiana University, Bloomington, and is Professor of Mathematics and Adjunct Professor of Computer Science, Informatics, Linguistics, and Philosophy. He is an editor of the *Journal of Logic, Language, and Information*, and of *Logical Methods in Computer Science*. He is Steering Committee Chair of the North American Summer School in Logic, Language, and Information.

Sarah Moss is Assistant Professor of Philosophy at the University of Michigan. Her current research applies tools from the philosophy of language to problems in metaphysics and formal epistemology. She has also published in formal semantics and pragmatics.

Allyson Mount is Assistant Professor of Philosophy at Keene State College in Keene, New Hampshire. Her research interests include context sensitivity and the role of speaker intentions, and she has published work on indexicals and demonstratives.

Ram Neta is Associate Professor of Philosophy at the University of North Carolina at Chapel Hill. He writes on various topics in epistemology. He is editor (with Duncan Pritchard) of *Arguing About Knowledge* (Routledge, 2009).

Bernhard Nickel is Assistant Professor of Philosophy at Harvard University. His research interests are in the philosophy of language, mind, and science, focusing especially on their interrelations.

Daniel Nolan is Professor of Philosophy at the Australian National University. He is the author of *Topics in the Philosophy of Possible Worlds* (Routledge, 2002) and *David Lewis* (Acumen, 2005).

Christopher Peacocke is Professor of Philosophy at Columbia University, and holds the Wollheim Chair of Philosophy at University College, London. He works on intentionality, understanding, metaphysics, epistemology, the philosophy of mind, and the relations between all of the preceding. His books include *Sense and Content* (Oxford University Press, 1983), *A Study of Concepts* (MIT Press, 1992), *Being Known* (Oxford, 1999), *The Realm of Reason* (Oxford University Press, 2004) and *Truly Understood* (Oxford University Press, 2008). He is currently working on a book on the self and the first person. He is a Fellow of the British Academy and of the American Academy of Arts and Sciences.

Francis Jeffry Pelletier is a Visiting Professor of Philosophy at the University of Alberta. He is an Emeritus Canada Research Chair in Cognitive Science and Emeritus Professor of Philosophy and of Linguistics at Simon Fraser University. Within the areas of philosophy of language and logic he has published work on generics, mass terms, and semantic compositionality. Additionally, he has published in the areas of automated theorem proving, artificial intelligence, cognitive science, ancient Greek philosophy, and history of logic.

Michael Potter is Professor of Logic at Cambridge University. His books include *Reason's Nearest Kin* (Oxford University Press, 2000), *Set Theory and its Philosophy* (Oxford University Press, 2004), and *Wittgenstein's Notes on Logic* (Oxford University Press, 2009). He has jointly edited *Mathematical Knowledge* (Oxford University Press, 2007) and *The Cambridge Companion to Frege* (Cambridge University Press, 2010).

François Recanati is a research fellow at CNRS and a professorial fellow at EHESS and the University of St Andrews, he has taught in several major universities around the world. His recent publications include *Literal Meaning* (Cambridge University Press, 2004), *Perspectival Thought* (Oxford University Press, 2007) and *Truth-Conditional*

Pragmatics (Oxford University Press, 2010). He is the Director of Institut Jean-Nicod (Ecole Normale Supérieure, Paris).

Teresa Robertson is Associate Professor of Philosophy at the University of California, Santa Barbara. Her primary interests are in topics that fall into both the areas of metaphysics and philosophy of language. She has published articles on origin essentialism, the sorites paradox, and reference to kinds.

Gillian Russell is an Associate Professor in the Philosophy Department at Washington University in St. Louis. She is the author of *Truth in Virtue of Meaning: A defence of the analytic/synthetic distinction* (Oxford University Press, 2008).

Sarah Sawyer is Senior Lecturer at the University of Sussex. Her research interests are based around the nature of and the connections between thought, language, and knowledge. Her published work primarily concerns content externalism, justification, fiction, and singular thought. She is also the editor of *New Waves in Philosophy of Language* (2010).

Barry Schein is Associate Professor of Linguistics in the Department of Linguistics at the University of Southern California. He has published on plurals and event semantics, and his current work is on the syntax and semantics of conjunction.

Mark Schroeder is Associate Professor of Philosophy at the University of Southern California. He works in ethics, practical reason, epistemology, and the philosophy of language, and is author of *Slaves of the Passions* (Oxford University Press, 2007), *Being For: Evaluating the Semantic Program of Expressivism* (Oxford University Press, 2008), and *Noncognitivism in Ethics* (Routledge, 2010).

Gila Sher is a Professor of Philosophy at the University of California, San Diego. She is the author of *The Bounds of Logic: A Generalized Viewpoint* (MIT Press, 1991). Her current research centers on logic, truth, and knowledge.

Sanford Shieh is Associate Professor of Philosophy at Wesleyan University. His research interests include philosophy of logic, history of analytic philosophy, philosophy of mathematics, and philosophy of language.

David Shier is Associate Professor of Philosophy at Washington State University. His published work is primarily on propositional attitudes, rationality, and the history of analytic philosophy.

Nicholas Smith is Senior Lecturer in Philosophy at the University of Sydney. He is the author of *Vagueness and Degrees of Truth* (Oxford University Press, 2008) and *Logic: The Laws of Truth* (Princeton University Press, 2012).

Scott Soames is Director of the School of Philosophy at the University of Southern California. His books include *What is Meaning?*, *Philosophy of Language*, *Reference and Description*, *Philosophical Analysis in the Twentieth Century*, *Beyond Rigidity*, *Understanding Truth*, and *Philosophical Essays*.

Robert J. Stainton is Professor of Philosophy at The University of Western Ontario. Recently, his research has focused on philosophical foundations of linguistics, history of philosophy of language, and pragmatic impairments.

Jason Stanley is Professor of Philosophy at Rutgers University, New Brunswick. He is the author of *Knowledge and Practical Interests* (Oxford University Press, 2005), *Language in Context* (Oxford University Press, 2007), and *Know How* (Oxford University Press, 2011).

Stephen Stich is Board of Governors Professor of Philosophy and Cognitive Science at Rutgers University and Honorary Professor of Philosophy at the University of Sheffield. He is a member of the American Academy of Arts and Sciences, a recipient of the Jean Nicod Prize, and the first recipient of the Gittler Award for Outstanding Scholarly Contribution in the Philosophy of the Social Sciences. His most recent book is *Collected Essays, vol. 1, Mind and Language* (Oxford University Press, 2010).

Martin Stokhof is Professor of Philosophy of Language at the Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam. He has published on various topics in semantics and pragmatics (such as questions and dynamic semantics), and on the philosophy of Wittgenstein. His current research interests focus on issues in the philosophy of linguistics.

Eric Swanson is Assistant Professor of Philosophy at University of Michigan, Ann Arbor. Much of his recent work is on the language and logic of causation, the language of subjective uncertainty, and extensions of supervaluationism.

Zoltán Gendler Szabó is A Professor of Philosophy at Yale University. His primary research focuses on the philosophy of language. He has written on compositionality, descriptions, context-sensitivity, tense, aspect, modality, and propositional attitudes.

Richmond Thomason is Professor of Philosophy, Linguistics, and Computer Science at the University of Michigan. He has edited Richard Montague's philosophical and linguistic works after Montague's untimely death in 1971, and his interest in formal semantics has persisted since that time.

J. Robert G. Williams is Professor in Theoretical Philosophy at the University of Leeds. His published work is primarily on topics in the philosophy of logic and language and metaphysics.

Seth Yalcin is an Assistant Professor of Philosophy at the University of California, Berkeley, where he is also a member of the Group in Logic and the Methodology of Science. His work usually has something to do with semantics, its foundations, and its interactions with metaphysics and epistemology.

ACKNOWLEDGMENTS

A volume of this size doesn't happen without substantial help, over a long period of time, from a large number of talented and committed people. So thank you, first of all, to our contributors. Our list of authors ended up looking like a *Who's Who* of contemporary philosophy of language and we are grateful to everyone who agreed to be a part of this project. But thanks also to the small army of philosophers who helped to referee papers and provide feedback for the authors. This volume could not have happened without their professionalism and selflessness. We are grateful to Ashley Atkins, a graduate student in the philosophy department at Princeton, for complinig the index.

We would also like to thank our editors Kate Ahl, and later Andy Beck and Mike Andrews at Routledge for their patience and for giving us the time to produce the book that we wanted.

Delia Graff Fara in particular acknowledges an enormous debt to Gillian Russell for taking on many lions' shares worth of work on this project. Her limitless enthusiasm, industriousness, diligence, as well as kindness are inestimable.

Gillian Russell would like to thank her partner, Robert Miller, for all his support. And also Meshuggah's Coffee on the Loop, where much of her work on the *Companion* was done. It's still her favorite.

INTRODUCTION

Gillian Russell

1

Do we ever make progress in answering philosophical questions? If we narrow our focus to the philosophy of language it is apparent that we do. Even acknowledging the difficulties in measuring such things, it's clear that more progress has been made in the last century or so than was made in any previous century, and moreover that there is reason to expect—if the present culture of training and research can endure—that our students will be able to do a better job than we.

We are thriving at least in part because of the tools and data we have received from two other thriving disciplines—both logic and linguistics came into their own in the 20th century. Thanks to logic we have been able to draw and consistently maintain distinctions that were previously beyond us and consider theories that we used to be unable to formulate. Thanks to linguistics we have learned about new phenomena and generalisations about those phenomena, and been exposed to data that has helped us to rule out or refine some of our own theories.

Readers who are new to the subject will want to ask what the philosophy of language is exactly and how it differs from logic or linguistics. Such a reader will probably be frustrated to discover that even researchers who have devoted their lives to the area are often unsure how to answer such questions and perhaps aren't too concerned by this. In practice, hesitation in defining the boundaries of one's discipline may present little obstacle to being a successful practitioner. Yet one way to introduce the philosophy of language to someone to whom it is entirely novel is to give examples of the questions with which it is concerned: what is meaning? what is truth? what is reference? How are meaning and reference related? How does linguistic communication work? All these questions are much more general than something that would be studied in a foreign language class, or even a class on the anthropology of language. But there is some danger that these questions sound so abstract and general as to give little idea what a philosopher of language actually does. Moreover, anyone who has actually studied the subject knows that it also involves questions of much greater specificity, such as: what is the meaning of the word *the?* What are the truth-conditions of subjunctive conditionals? And: how is anaphoric reference determined? And if we put it this way the subject is in danger of sounding so concrete as to be dull, pedantic, and perhaps worst of all, like we are attempting to do linguistics without getting proper training in the subject first. That might have been acceptable before there *was* a developed science of linguistics, but it would be jejune now.

We can perhaps do better by exploring an example. Take our first question: what is meaning? We normally think that sentences have meanings—they say something—and that sentences in two different languages, and even two different sentences in the same language, may have the same meaning. But just what kind of thing *is* a meaning? Some observations can help to restrict the set of possible answers: what sentences say—their meanings—are often the kinds of thing that can be true or false. The meaning of a whole sentence is a function of the meanings of its parts. Even sentences that no-one has ever said or thought of may have meanings. Some sentences seem to be useable with different meanings on different occasions.

That might not seem like much to go on—and I don't wish to suggest that any of these claims is unassailable if we discover good reasons to reject them—but this is somewhere to start. It tells us, for example, that meanings must be the kinds of things that can have parts, that such parts must be composable into something that can be either true or false, and that there is no simple 1–1 relationship between sentences and their meanings. There is much information that could be directly or indirectly useful in taking such a miniature project further. Some of it could come from psychology: meanings are the kinds of things that speakers of a language are sensitive to, so they must somehow be the kinds of things that human beings *can* be sensitive to. Psychology—including things like studies of language learning, or linguistic processing by speakers—can tell us more about how that can happen. Other information comes from linguistics, especially from the study of specific features of languages. Meanings can be complex, and the ways in which meanings compose into more complex meanings is quite various. We would not have the conceptual tools necessary for formulating appropriate theories without the help of mathematicians and logicians. Recent research into the psychology of very basic mathematical abilities in humans and animals has suggested that language was an important tool in the development of human mathematical capacities, but, it is also quite clear from looking at contemporary philosophy of language that our theories of language would be miserably poor without mathematics. Taking all this scientific information and all these mathematical tools, and coming up with theories about what meanings are and reasons to accept some of those theories over others, is one of the things that philosophers of language do.

2

There are lots of features of natural languages of which their native speakers are ignorant. That is perhaps a contentious way to put it, but the situation is roughly this: you can be a competent speaker of English without being able to produce an explicit list of the syntactic markers for context-sensitivity, the rules for turning active constructions into passives, or for determining the referents of anaphora. (It gets worse. If Tim Williamson's *Philosophy of Philosophy* is to be believed, you can be a competent speaker of English and not know that all vixens are vixens.) Competent English speakers often *are* nonetheless able—as they wander about the world, having conversations and reading each others postcards—to judge accurately the truth-conditions of sentences containing context-sensitive expressions, reliably turn actives into passives, and determine the referents of anaphora. It is just that when asked to come up with a general theory of how such things are done, if they give an answer at all it will usually be wrong. We might say that for humans speaking their native languages, linguistic processing—much like visual object processing—is mostly transparent and automatic. We don't so much see

it happening, or intentionally make it happen, as simply find ourselves presented with the results. Once we acknowledge this it becomes unsurprising that there is much to be discovered about how language works, including things that no speaker knows or has ever known.

Still, the myth that competent speakers know all there is to know about their languages has historically been a barrier to progress, because it has suggested that there is nothing about a living language that some living person doesn't already know. One way in which linguistics has greatly aided the philosophy of language is by making it increasingly obvious that there is a large gap between what every fluent speaker knows, and what the experts know. In doing so they have made it clear that more substantive research into language is possible.

Yet even today, the misconception of full speaker knowledge survives where it has not been wiped out, and unfortunately that includes many areas of philosophy. Where it survives it has a tendency to underwrite bad work on language, and undermine the reception of the good. It underwrites the bad by suggesting that if one is a native speaker (and perhaps in addition a philosopher of science, or an ethicist, or a metaphysician) and one proposes a theory of how some part of the language works and can see no problems with it, then there *are* no problems with it. A naive philosopher might simply conclude that there is nothing substantive or serious to be said against his claims about meanings because he has a hard time envisioning what serious or substantive work there could be. It undermines the reception of good work because having a favourite answer to a problem and thinking that there can be no serious or substantive work that contradicts it tends to undermine one's motivation to read the literature.

For example, there are areas in philosophy—including value theory, vagueness, the philosophy of arithmetic, and the philosophy of ordinary objects—where we need both an account of what sentences about the domain of study are claiming, and an account of what exists and is happening within the domain itself. We need some philosophy of language and we need some metaphysics. Sometimes one's theory of what exists and is happening in the domain itself can be made more satisfying by a novel account of the semantics of the sentences about the domain. Ethicists might suggest that value-ascribing sentences don't really make assertions, and thereby free themselves from the requirement of positing certain odd kinds of object. Philosophers of mathematics might claim that numerals are not really referring expressions, and that sentences employing them don't require the existence of numbers to be true, and metaphysicians may find it simplifies their metaphysical theories if they maintain that some locution in English is systematically ambiguous, or context-sensitive, or creates an opaque context. In the most egregious cases, this is treated as if it can be done "for free." The fact that it results in an improved metaphysical theory, and *the author* (who is, after all, a competent speaker) sees no problems with the linguistic part of the theory is taken as grounds enough—as if there could be no independent, substantial linguistic research that bore on the question of whether the linguistic part of the theory is true.

That's a mistake, and to ignore the research that has been done is to ignore significant evidence that is relevant to whether or not one's overall picture is correct. It is possible that non-philosophers of language will find the extra tests that their combined theories have to pass onerous. But from the disinterested perspective of knowledge gathering, a rich new source of evidence is to be welcomed—it gives us a new tool for telling the good theories from the bad.

3

The philosophy of language can itself be of great use to the rest of philosophy, but there is much more to philosophy than just the philosophy of language. There are genuinely hard questions in ethics and metaphysics, in epistemology and the philosophy of science, that would remain unsolved even if all questions in the philosophy of language were answered. It is as much a mistake to think that the only substantive philosophy is philosophy of language as it is to think that nothing concerning language is all that substantive (and woe betide those philosophers who have done both). Still, the claim that I want to make here is stronger than many readers will like: I claim that if you want to be a good philosopher today then you ought to study a lot of philosophy of language.

The argument for this is both simple and a bit dull. Arguments are the primary tools and targets of philosophers. When we assess an argument we consider whether the premises are true, and whether the premises support the conclusion. But the answer to such questions depends on the meaning of the sentences in the arguments, and so it is understandably common for opposing philosophical parties to dispute claims about meaning. Hence philosophers must often evaluate claims about meaning. Meaning controversies include whether *knows* is a context-sensitive expression, whether value attributions assert propositions and whether numerals are singular terms. Such questions are not trivial ones, which any fluent speaker is in prime position to answer accurately with relatively little thought. Nor can the answers be looked up in the index of the average linguistics textbook. Facility in the philosophy of language makes you better able to imagine, distinguish and evaluate the different possible answers, and as a result less likely to make a mistake which will hobble your account in another areas.

It is just possible that philosophers in other areas might find this situation trying. Perhaps they became epistemologists, ethicists or philosophers of mind etc . . . because they wanted to do epistemology, ethics or philosophy of mind, where they did not expect that this involved ploughing through a densely written literature on quantifiers or gradable adjectives. These other parts of philosophy are no doubt abstract and technical disciplines themselves, but to make the situation worse, all the input from logic and linguistics has made some philosophy of language unusually technical and specialised too, and thus somewhat inaccessible.

So what is a conscientious philosopher to do? Suppose you are already embroiled in the debate in some area, and you begin to suspect that you ought to know more about the semantics of questions, or mass terms or subjunctive conditionals. Perhaps this is because other people in the debate seem to know a lot about it already. Perhaps it is because when an opponent makes a certain claim you have an inchoate hunch that their theory clashes with the true semantics and you wish you knew what that was, so that you could formulate your objection neatly and persuasively. Or perhaps—and maybe this is the really exciting possibility—perhaps you are suddenly struck with a totally original idea and it involves claiming that such and such a claim is really just equivalent to another, less confusing claim, and you want to check and see whether such a thing would be semantically feasible. You pop next door to see your friendly, neighbourhood philosopher of language, but she is too busy with the metaphysicians. What should you do? Fortunately, this is just the kind of emergency that this *Routledge Companion to the Philosophy of Language* is designed to handle.

Contents

This *Companion* is intended to make the literature on many topics in the philosophy of language accessible to a variety of audiences: those who are studying the subject for the first time, those who are already philosophers working in other areas, and even professional philosophers of language who might want a survey of contemporary work in some unfamiliar sub-area. For each chapter, we picked a top researcher and asked them to write a short, interesting introduction to their topic. Many of these introductions are quite opinionated, but where they take a side on a controversy you should find that they also give you a good sense of where to look for the alternative approaches, and what you might put on your reading list if you wanted to know more.

Section I presents work on some of the core ideas in the philosophy of language, including extensions, intensions, character, presupposition, implicature, communication, logical form, and context-sensitivity.

Section II then focuses on approaches to some of the key topics that fall under the heading of 'semantics': reference, propositions, concepts, and some rival approaches to giving a theory of meaning, including possible world semantics, event-based semantics, and one of the more high-profile and intriguing newcomers, dynamic semantics.

Section III groups together entries focusing on particular kinds of expression, ranging from words like names and adjectives, to complex expressions like a conditionals, descriptions and questions. Some of these topics—such as descriptions—have already played an important role in philosophy, and in fact we might say that it was philosophers who kickstarted the linguistics literature. Others—such as questions—have come to play a role in philosophy only more recently.

The methodology of philosophy—a topic in the philosophy of philosophy—has attracted significant interest in recent years, and Section IV of this book could be regarded as a new contribution to that field. Indeed, it could almost have been published as a stand alone book containing six essays on the methodology of the philosophy of language. With other parts of the *Companion* we were usually asking researchers to write introductions to literatures that we knew already existed, but in this one we wondered what people would have to say, and commissioned the articles to find out: we wanted a new literature on the topic, one that would contribute to the "methodological awareness" that Timothy Williamson recommends in his *Philosophy of Philosophy* and which could serve as a series of case studies for people writing about the methodology of philosophy more generally. Hence we asked Stephen Stich and Edouard Machery to write about the use of experiment in the philosophy of language, Robert Stainton to cover the role of psychology, Larry Moss to talk about mathematical methods, Sarah Moss to introduce the uses of linguistics, Michael Devitt to talk about intuitions, and finally Martin Stokhof to discuss the role of artificial languages.

Section V turns the focus to logic. The role that logic plays in the study of language is itself a controversial, and the section opens with John P. Burgess' cautionary sermon "Model Theory: what it is and isn't," which warns of the confusions that stem from confusing model theory with semantics. Subsequent articles introduce and summarise research both on topics that have traditionally been important to the philosophy of language, such as quantification and many-valued logics, and on work in logic that has risen to prominence recently, including two dimensional and dynamic logics.

Section VI focuses on the uses that philosophy of language has been put to in philosophy more broadly. It includes essays on the use of philosophy of language in epistemology

and metaethics, and we asked philosophers in metaphysics and normative ethics to write essays on language *and* some particular topic in their area, giving us, for example, the team of Luvell Anderson, Rae Langton, and Sally Haslanger on language and race, and Eric Swanson on language and causation.

Contemporary philosophy of language is not an especially historically-oriented part of philosophy. The historians I know bemoan this, and with some grounds. A standard introductory course in our area starts in the late 19th century with Frege—not with the pre-Socratics—and just last week I was talking to a philosopher of language who suggested that it was a mistake to make students read original texts (such as *Kripke and Burge)* and that we would be better off using contemporary textbooks, much as a logician or a linguist would. The argument appealed to the thought that the big ideas are rarely presented in their clearest form when they are first discovered, and it is much easier to learn about them from cleaned-up, contemporary versions. Be that as it may, the history of philosophy of language is a serious area of study in its own right and ignorance is not something to be proud of. Historical ignorance, in particular, may lead to the repetition of old mistakes and this means that even the most contemporary of philosophers of language has good reason to study her subject's history. Our final section, Section VII attempts to provide some. We have gone beyond the familiar story that starts with Frege, including sections on ancient, medieval and modern philosophy as well as a chapter on the history of generative grammar. The latter is likely to make fascinating reading for younger philosophers, even if older readers are likely to be horrified to find events they lived through appearing in the history section. Despite our efforts, the philosophers who wrote these sections have stressed how much more there is to be said (especially when we attempted to keep them to word limits). Nonetheless, much of what is here will be new to the majority of readers and each chapter includes pointers about where to go if you would like to learn more.

On our cover image

Finally, the cover image for this book is a photograph of five surgical tools, and we promised the ever-patient Andrew Beck—our editor at Routledge—that we would explain the choice. Typical jackets for books on language feature famous paintings of people talking or moderately abstract images containing two heads. Possibly in pastel colours. Our image is less fluffy. We picked it because the philosophy of language is much less *fluffy* than outsiders normally suppose. We pride ourselves on clear, precise, both empirically and mathematically informed analysis of subjects that at first glance appear to be messy and organic. Surgical tools that are designed for use on the human body seemed like an apt metaphor. In addition, we want to emphasise the extent to which philosophy of language can itself be a precision tool—something that philosophers in other areas can exploit in their own work. Philosophical questions can be difficult and confusing, muddled and, at first, rather messy. We think of this *Companion* as offering you tools designed for such work. *Achtung!*—they're sharp.

Section I

CORE TOPICS

1.1

EXTENSION, INTENSION, CHARACTER, AND BEYOND

David Braun

This article explains some of the technical terms commonly used in semantic theory. One such term is *extension*. The notion of an extension is a generalization of the notion of reference. Proper names refer to objects, and a proper name's extension is the object to which it refers. For instance, the extension of 'Barack Obama' is Barack Obama, and the extension of 'Los Angeles' is Los Angeles. The extension of a predicate is the set of objects to which it correctly applies. For instance, the extension of the common noun 'dog' is the set of all dogs, the extension of the verb 'run' is the set of all things that run, and the extension of the adjective 'happy' is the set of all happy things. Binary predicates correctly apply to pairs of things, and so the extension of a binary predicate is a set of ordered pairs. For example, the extension of the binary predicate 'kiss' is the set of pairs of objects $<x, y>$ such that x kisses y. The extension of a ternary predicate (such as 'give') is a set of ordered triples. Generally, the extension of an n-place predicate (where n is greater than 1) is a set of n-tuples.

Complex expressions also have extensions. For instance, the extension of 'kisses Barack Obama' is the set of objects to which this phrase correctly applies, namely the set of objects that kiss Barack Obama. The extension of this complex expression is completely determined by the extensions of the simple expressions in it: for an individual z is a member of the extension of 'kiss Barack Obama' if and only if there is a pair of objects $<x, y>$ such that (i) $<x, y>$ is a member of the extension of 'kiss', (ii) y is identical with the extension of 'Barack Obama' (that is, y is identical with Barack Obama), and (iii) z is identical with x. Declarative sentences are also complex expressions. If we assume that they have extensions, and we also assume that the extension of every complex expression is completely determined by the extensions of the simple expressions in it, then we are naturally led to the conclusion that the extension of a declarative sentence is its *truth-value* (either *truth*, if the sentence is true, or *falsehood*, if the sentence is false). For example, the sentence 'Barack Obama runs' is true if and only if Barack Obama runs. But the extension of 'Barack Obama' is Barack Obama and the extension of 'runs' is the set of things that run. So the extension of the sentence 'Barack Obama runs' is truth if and only if the extension of 'Barack Obama' is a member of the extension of 'runs'. Therefore, the truth-value of the sentence 'Barack Obama runs' is

completely determined by the extensions of the words in that sentence. Similar points hold for many other simple sentences.

Extensions are important to semantics, but they are not plausible candidates for the *meanings* of expressions, for expressions that have the same extension (*co-extensive* expressions) can differ in meaning. For example, the predicate 'renate' is co-extensive with the predicate 'cordate', and yet the terms differ in meaning, for 'renate' means *animal with a kidney* whereas cordate means *animal with a heart*. Every true sentence has the same extension (truth), and every false sentence has the same extension (falsehood), but obviously there are pairs of true sentences, and pairs of false sentences, that differ in meaning: consider, for example, the true sentences 'Some dogs live in Pittsburgh' and 'All cats are mammals' and the false sentences 'Some cow is purple' and 'Every frog is a mammal'. Some pairs of co-extensive (co-referring) proper names, such as 'George Orwell' and 'Eric Blair', also seem to differ in meaning, for the sentence 'George Orwell is George Orwell' seems to differ in meaning from 'George Orwell is Eric Blair'.

We can sum up the problems with identifying meanings with extensions by saying that meanings are *finer-grained* semantic entities than extensions. Generally, one sort of semantic entity *X* is finer-grained than another semantic entity *Y* if two expressions that have the same *Y* can have different *X*'s.

If meanings are not extensions, then what are they? Many traditional theories say that the meanings of expressions are individuals and attributes, and complex entities made up of these. Typically, these views say that every predicate *expresses* one or more attributes, and the meaning of a predicate is either the attribute it expresses or the set of all attributes it expresses. For instance, on some versions of this theory, 'red' expresses the property of being red, while 'human' expresses the properties of being rational and being an animal. Therefore, the meaning of 'red' is the property of being red, and the meaning of 'human' is the set consisting of the properties of being rational and being an animal. The meaning of 'renate' is the property of being an animal with a kidney and the meaning of 'cordate' is the property of being an animal with a heart. The latter two properties are distinct, and so the predicates 'renate' and 'cordate' differ in meaning. The meanings of these predicates determine their extensions. For instance, the extension of 'renate' is the set of all things that have the property of being a renate, and so the latter property determines the extension. Similarly for other predicates: the meaning of 'kiss' is the binary relation of kissing, and this relation determines the term's extension: namely the set of all pairs $<x, y>$ such that x bears the relation of kissing to y.

On some versions of this traditional theory, the meaning of a proper name is the object to which it refers, and so the names 'George Orwell' and 'Eric Blair' have the same meaning. If the meaning of a proper name is its referent, then the meaning of a name on this view obviously determines its extension, for they are one and the same thing. On other versions of the traditional theory, the meaning of a proper name is the same as that of some definite description: for instance, the meaning of 'George Orwell' is the same as that of 'the author of *1984*'. The meaning of a definite description is, roughly speaking, a property that at most one object can bear. Thus the meaning of 'George Orwell' might be the property of being the (sole) author of *1984*, whereas the meaning of 'Eric Blair' might be some other property that only Orwell/Blair has. If the meaning of a proper name is a property that at most one object can satisfy, then the meaning determines the referent by determining a unique object.

Complex expressions, such as 'kisses Barack Obama', also have meanings. On many traditional views, the meaning of a complex expression *E* is a structured entity, whose

structure reflects the grammar of *E* and whose constituents are the meanings of the words in *E*. For instance, the meaning of 'kisses Barack Obama' is something like a sequence of the meanings of the words in the phrase ('kiss' and 'Barack Obama'), in the order in which the words appear in the phrase, or a tree-like structure whose ultimate constituents are the meanings of the words in the phrase. On such views, the meaning of a complex phrase determines its extension. For instance, the meaning of 'kisses Barack Obama' contains as constituents the meanings of 'kiss' and 'Barack Obama'. The meaning of 'kiss' is the relation of kissing, and (as we saw above) this determines the extension of 'kiss'. The meaning of 'Barack Obama' is either Barack Obama himself or some property that he has uniquely; either way, the meaning of 'Barack Obama' determines the extension of the name. Finally, as we saw before, the extensions of 'kiss' and 'Barack Obama' determine the extension of the phrase 'kisses Barack Obama'.

The meaning of a declarative sentence is traditionally called a *proposition*. Declarative sentences are said to *express* propositions. On many traditional views, propositions are complex structured entities: if sentence *S* expresses proposition *P*, then *P* has a structure that reflects the grammar of *S*, and the ultimate constituents of *P* are the meanings of the words in *S*. For instance, the proposition expressed by 'Barack Obama runs' has the meanings of 'Barack Obama' and 'runs' as constituents. Propositions that have this sort of structure are often called *structured propositions* or *Russellian propositions*, after Bertrand Russell (1903), who uses them extensively in his semantic theories. (See also Kaplan 1989 and Salmon 1986.) A proposition that has an individual as a constituent is often called a *singular proposition*. The proposition that a sentence expresses determines the sentence's extension (its truth-value). For instance, the meaning of 'Barack Obama runs' has the meanings of 'Barack Obama' and 'runs' as constituents. These meanings determine the extensions of 'Barack Obama' and 'runs', and as we saw earlier, the extensions of these terms determine the truth-value of the sentence.

Many traditional theories of meaning say that people can bear various cognitive and linguistic relations to propositions. For instance, people can assert, deny, believe, or doubt the proposition that Barack Obama runs. Furthermore, propositions are the referents, and extensions, of 'that'-clauses: for example, the extension of 'that Barack Obama runs' is the proposition that Barack Obama runs. Therefore, the sentence 'Clinton believes that Barack Obama runs' attributes to the referent of 'Clinton' belief in the proposition that Barack Obama runs.

Some theorists use alternative terminology for similar notions. Some use 'denotation' for (roughly) the extension of a term, and 'denote' for the relation between an expression and its denotation (Mill 1872, Book 1, Chapter 2, Section 5; Russell 1903). Others use 'designate' and 'designation' for similar extensional notions. Some use 'connotation' for meaning, and 'connote' for the relation between an expression and its connotation (Mill 1872). 'Comprehension' is an older term (rarely used now) for the meaning of a predicate (Arnauld 1662/1964, 51). *Intension* is a more common term for meaning or connotation (Hamilton 1860, Volume 2, 101; Carnap 1956). Semantic theories that describe only the extensions of expressions are often called *extensional semantic theories*. Semantic theories that attempt to describe the meanings of expressions are often called *intensional semantic theories*.

A more recent, twentieth-century proposal for describing the meanings of expressions relies on the idea that the extension of a term often depends crucially on the way the world is. For instance, the predicates 'renate' and 'cordate' are in fact co-extensive, but there could have been animals with kidneys but no hearts, or animals with hearts

but no kidneys, and if either had been the case, then the extension of 'renate' would have been different from the extension of 'cordate'. We can use this idea to try to distinguish between the meanings of 'renate' and 'cordate', in the following way. Consider all of the different ways the world might have been, and call these ways *possible worlds*. The extension of 'renate' *at a possible world* is the set of things that are renates *in that world*. Thus the extension of 'renate' varies from possible world to possible world. Now we can define a *function* from possible worlds to extensions as follows: Given any possible world as an input (or *argument*), this function yields as its output (or *value*) the set of things that are renates in that possible world. This function is the *possible-worlds intension* of 'renate', and according to this theory, the meaning of 'renate' is identical with this function. (Most theorists who accept this theory call these functions simply 'intensions', leaving out 'possible-worlds'. I will say more about this terminology later.) Similarly, the extension of 'cordate' varies from possible world to possible world, and the possible-worlds intension of 'cordate' is a function whose value at any possible world is the set of things that are cordates in that world. Since it is possible that some cordates are not renates, the extensions of 'renate' and 'cordate' differ at some possible world, and so the possible-worlds intensions of these terms are different functions. If the only meanings of these terms are their possible-worlds intensions, then they differ in meaning.

We can extend this idea to other expressions. The predicate 'kiss' has an extension at each possible world (the set of ordered pairs in which the first member kisses the second in that world), and the possible-worlds intension of 'kiss' is the function that yields the extension of 'kiss' at each possible world. Similarly for other predicates. The possible-worlds intension of a proper name is a function which, at every possible world, yields the extension of that name at that world. Suppose that the extension of a given proper name can vary from possible world to possible world. Then there may be worlds at which the extension of 'Eric Blair' is different from the extension of 'George Orwell'. If so, then the possible-worlds intensions of the names are different, and if meanings are possible-worlds intensions, then these names differ in meaning. The extension of a declarative sentence at a possible world is a truth-value, and its possible-worlds intension is a function that, at every world, yields the sentence's truth-value at that world. There are some worlds in which no dogs live in Pittsburgh, but all cats are mammals. In these worlds, the extension of 'Some dogs live in Pittsburgh' is falsehood and the extension of 'All cats are mammals' is truth. Therefore, these sentences have different possible-worlds intensions, and if their meanings are their possible-worlds intensions, then they differ in meaning.

The view that meanings are identical with possible-worlds intensions is sometimes called *possible-worlds semantics*. Many proponents of possible-worlds semantics identify the proposition expressed by a sentence *S* with its possible-worlds intension, but some instead identify the proposition expressed by *S* with the set of possible worlds in which *S* is true. These two ways of speaking about propositions are equivalent, on possible-worlds semantics. The possible-worlds intension of *S* divides up the set of possible worlds into two sets, the set of worlds in which *S* is true and the set of worlds in which *S* is false. Therefore, *S*'s possible-worlds intension determines the set of worlds in which *S* is true. But the set of worlds in which *S* is true also determines *S*'s intension, for if *W* is the set of worlds at which *S* is true, then the intension of *S* is the function that yields truth for every world in *W* and yields falsehood for every world not in *W*. Therefore the two ways of speaking of propositions are equivalent, according to possible-worlds semantics.

Traditionally, the term 'intension' has been used as a virtual synonym for 'meaning'. But most contemporary philosophers and linguists use the term 'intension' specifically

for functions from possible worlds to extensions; these are the functions that I called 'possible-worlds intensions' above. Some philosophers and linguists who use the term 'intension' in this modern way think that meanings really are identical with such functions (Carnap 1956, Lewis 1972, Montague 1974, Stalnaker 1984). Others who use 'intension' in this modern way accept that there are such functions but deny that meanings are identical with them (Kaplan 1989, Salmon 1986, Soames 1987). To avoid taking a stand on this issue, I shall continue to use the term 'possible-world intension' when I speak of functions from possible worlds to extensions, and I shall continue to use plain old 'intension' (without the modifier 'possible-world') in the traditional way, as a term that is, more or less, synonymous with 'meaning'. Readers should keep in mind that I am, in this respect, departing from standard contemporary usage.

Along with 'extension' and 'intension' come other related technical terms. An *extensional context* is a part of a complex expression in which substitution of co-extensive expressions does not change the extension of the complex expression. For instance, the underlined portion of the verb phrase 'kissed Eric Blair' is an extensional context in that verb phrase, because substituting a co-extensive expression, such as 'George Orwell', for 'Eric Blair' in this context does not change the extension of the verb phrase (namely, the set of things that kissed Blair/Orwell). Similarly, the underlined portion of 'Eric Blair kissed Mary' is an extensional context in this sentence, for if one substitutes a co-extensive expression for 'Eric Blair' (e.g., 'George Orwell'), then the extension (the truth-value) of the entire sentence remains the same. A *non-extensional context* is a part of a complex expression in which substitution of co-referring expressions may change the extension of the complex expression. For instance, the underlined portion of 'Necessarily, all renates have kidneys' is a non-extensional context in the sentence, for substituting the co-extensive expression 'cordate' for 'renate' yields a false sentence. Similarly, substituting 'cordate' for 'renate' in the sentence 'John believes that all renates have kidneys' may result in a change in the truth-value of this sentence, and so the underlined portion is a non-extensional context in the sentence. Sentences that contain non-extensional contexts are exceptions to the general rule that the extension of a complex expression is completely determined by the extensions of its parts. An *intensional context* is a part of a complex expression in which substitution of expressions with the same possible-worlds intension does not change the extension of the complex expression. For instance, the underlined portion of 'Necessarily, all renates have kidneys' is an intensional context in this sentence, for we can substitute any expression with the same intension as 'renate', such as 'animal with kidneys', and the truth-value of the whole sentence will remain the same. (Different theorists use the term 'intensional context' differently. By my definition, the underlined portion of 'Eric Blair kissed Mary' is both an extensional and an intensional context. But some theorists use the term 'intensional context' for all and only contexts that are non-extensional; they would not count this context as intensional.) A *hyperintensional context* is a part of a complex expression in which substitution of expressions with the same possible-worlds intension may change the extension of the complex expression. For example, the underlined portion of 'John believes that all animals with kidneys have kidneys' is a hyperintensional context in the sentence, for substituting 'renate' for 'animals with kidneys' may change the truth-value of the sentence.

Possible-worlds intensions resemble meanings in at least two respects. First, meanings determine extensions, and so do possible-worlds intensions. Second, two expressions with the same extension can differ in meaning, and likewise two expressions with the

same extension can differ in their possible-worlds intensions. But identifying meanings with possible-worlds intensions is problematic (Soames 1987). The sentences 'All dogs are dogs' and '2 + 2 = 4' are both true at all possible worlds, and so have the same possible-worlds intension, but seem to differ in meaning. The sentences 'Barack Obama exists and he is not an electron' and 'Barack Obama exists and 2 + 2 = 4' are both true at all possible worlds in which Obama exists, and both false at all worlds in which Obama does not exist, and so have the same truth-value at all worlds, and therefore have the same possible-worlds intension, but they differ in meaning. The predicates 'triangular' and 'trilateral' have the same extension at any possible world, and so they have the same possible-worlds intension, but they differ in meaning. Many philosophers think that the extension of a given proper name does not vary from world to world (see Kripke 1980 for arguments for this claim). On their view, proper names are *rigid designators*, that is, expressions whose extensions, or designations, are the same at all possible worlds. If proper names are rigid designators, then two proper names that have the same extension at the actual world, such as 'Eric Blair' and 'George Orwell', have the same extension at all possible worlds, and so have the same possible-worlds intension. Yet some such pairs of names seem to differ in meaning.

To deal with some of these difficulties in identifying meanings with possible-worlds intensions, some philosophers propose to identify meanings with *structured intensions* (Lewis 1972, Creswell 1985). A structured intension is a structured entity that contains possible-worlds intensions as constituents. The structured intension of a sentence *S* is an entity that contains the possible-worlds intensions of the words in *S*, within a structure that reflects the grammatical structure of *S*. The structured intension of 'All dogs are dogs' contains the possible-worlds intension of 'dog', while the structured intension of '2 + 2 = 4' contains the possible-worlds intension of '2' and does not contain the possible-worlds intension of 'dog'. Therefore, these two sentences have different structured intensions, even though they have the same (unstructured, possible-worlds) intension. So if meanings are structured intensions, then these sentences differ in meaning. Similarly, 'Barack Obama exists and is not an electron' has a different structured intension from 'Barack Obama exists and 2 + 2 = 4' because (for instance) the first contains the possible-worlds intension of 'electron' whereas the second does not. Complex predicate phrases, such as 'exists and is not an electron' and 'exists and is such that 2 + 2 = 4' also have different structured intensions. The expressions 'trilateral' and 'triangular' also have different structured intensions, if they are complex predicate phrases made up of 'tri' and 'angular', and 'tri' and 'lateral', respectively.

Structured intensions resemble the meanings that the earlier traditional theories attribute to complex expressions. On the traditional theory, the meaning of 'Barack Obama runs' is a complex entity that has as its constituents the meaning of 'Barack Obama' (which is either the man or a certain sort of attribute) and the meaning of 'runs' (the property of running). The structured intension of the sentence is similar, but has the possible-worlds intensions of 'Barack Obama' and 'runs' as constituents. However, some philosophers and linguists hold that attributes just are possible-worlds intensions. This identification is controversial, but if it is correct, then structured intensions come close to the meanings of the traditional theory. There remains one difference, however, for some traditional propositions contain individuals (such as Barack Obama) as constituents, whereas no structured intension does (though a structured intension can contain a possible-worlds intension whose value at every world is Barack Obama).

Co-referring proper names still present a serious apparent problem for both theories. Suppose that proper names are rigid designators. Then 'George Orwell' and 'Eric Blair' have the same possible-worlds intension. Therefore, the sentences 'George Orwell is George Orwell' and 'George Orwell is Eric Blair' have the same structured intension. Yet the sentences seem to differ in meaning. Traditional theories that say that the meaning of a proper name is its referent have a parallel problem, for these theories entail that the two names have the same meaning, and so entail that the two sentences have the same meaning. An advocate of structured intensions might try to evade this problem by denying that proper names are rigid designators; if 'George Orwell' and 'Eric Blair' differ in extension at some world, then they have different possible-worlds intensions, and so differ in meaning. A traditional theorist might deny that the meaning of a name is its referent, and instead hold that the meaning of a name is an attribute, and claim that the meanings of these names are distinct attributes. But most theorists reject these options, for reasons given by Kripke 1980 (see also Chapter 3.1 on Names in this volume).

Gottlob Frege (1892) distinguishes between an expression's *sense* and its *reference*. (He uses the German terms 'Sinn' and 'Bedeutung', respectively.) Frege's notion of reference is roughly the same as that of extension; his notion of sense corresponds with the notion of meaning we have discussed so far. But Frege's theory of sense seemingly entails that senses are finer-grained than both structured intensions and traditional meanings. Frege held that if a rational speaker can understand two sentences *S* and *S'*, and think that *S* is true and *S'* is false, then *S* and *S'* differ in sense. A rational speaker can understand 'George Orwell is George Orwell' and 'George Orwell is Eric Blair', and yet think that the first is true and the second is false. So Frege's theory entails that these sentences differ in sense. Therefore the Fregean senses of proper names are finer-grained than possible-worlds intensions, and the Fregean senses of proper names are also finer-grained than the meanings that certain traditional views attribute to names: namely those views that identify the meaning of a proper name with its referent. Parallel considerations seemingly show that the Fregean senses of predicates are finer-grained than possible-worlds intensions and attributes. Consider the sentences 'All pails are pails' and 'All pails are buckets'. The property of being a pail is the same as the property of being a bucket, and the possible-worlds intension of 'pail' is the same as that of 'bucket'. Yet it seems that a rational person could understand both of these sentences and think that the first is true and the second is false. If this is right, then Fregean senses cannot be identified with possible-worlds intensions or the attributes of traditional theories of meaning. Philosophers are divided on the plausibility and viability of a theory of fine-grained Fregean senses. See Kripke (1980) and Salmon (1986) for discussion.

We need even further distinctions, further sorts of semantic entities, and further semantic vocabulary, to describe a common phenomenon that we have ignored so far: namely that a single sentence can "say" different things in different contexts. Imagine that John and Mary both utter the sentence 'I run'. This sentence is not ambiguous (or at least it is not ambiguous in the way that 'bank' and 'John is standing next to a bank' are). Nevertheless, John and Mary assert different propositions, for John says that he runs while Mary says that she runs (and John does not say that Mary runs and Mary does not say that John runs). What they say may even differ in truth-value, for it may be true that John runs, while false that Mary does. In view of this, it is natural to think that the sentence 'I run' has a single meaning, but that it expresses different propositions on different occasions, depending on who utters it. Something similar happens with sentences containing 'here', 'today', 'yesterday', 'now' 'he', 'she', 'this', and 'that'. Two people who

utter a single sentence containing one of these words may very well end up asserting different propositions.

We can systematize these observations in the following way (as Kaplan 1989 does). Let us suppose that words and sentences have two kinds of meaning, *character* and *content*. The sentence 'I run' has a single character. The character of this sentence is the meaning that all utterances of it have in common. But this sentence has different *contents* in different *contexts*. The content of the sentence in a context is the proposition that it expresses in that context. The content of 'I run' in John's context is the proposition that John runs. The content of that same sentence in Mary's context is the proposition that Mary runs. The character of the sentence determines what its content will be in a context. Similar points can be made about single words. The word 'I' has a single character, but varies in content from context to context, depending on who the agent of the context is. When an expression varies in its content from context to context, we call it 'context-sensitive'. Some expressions, such as 'runs', have the same context in all contexts, and so are context-*in*sensitive.

The meanings that we discussed earlier in this chapter should be identified with the *contents* of this new view that distinguishes between character and content. The claim that there are two kinds of meaning, character and content, is compatible with various views about the nature of contents. If contents are like the meanings of some traditional theories, then the content of 'I run' in John's context is a structured proposition that contains John himself and the property of running as constituents. If contents are possible-worlds intensions, then the content of 'I run' in John's context is a function from possible worlds to truth-values whose value at any world is truth iff John runs at that world. Similarly, the content of 'I' in John's context is John himself, if contents are like the meanings of certain traditional theories; if contents are possible-worlds intensions, then the content of 'I' in John's context is a function whose value at every world is John. From here on, I will (for the sake of simplicity) assume a traditional theory of meaning on which contents are individuals, attributes, propositions, and complexes of such entities.

The character of an expression is usually identified with a function from contexts to contents, whose value at any context is the content of that expression in that context. For instance, the character of 'I' is a function on contexts whose value at any context is the content of 'I' in that context: namely the agent of the context. Given John's context, this character yields John himself, and given Mary's context, it yields Mary herself. The character of 'runs' is a constant function whose value, at every context, is the property of running. The character of 'I run' is a function whose value at every context is the singular proposition that *A* runs, where *A* is the agent of the context.

Typical theories of character (such as that of Kaplan 1989) assume that every context has its own agent, time, location, and possible world, and also assume that the agent of a context exists in the world of that context, and is located at the time and place of that context in that world. A sentence is true at a context C iff its content in C is true in the world of C. For instance, the sentence 'I exist' is true in a context C iff the content of that sentence in C is true in the world of C. Therefore 'I exist' is true in every context, for the agent of a context always exists in the world of the context. Similarly, 'I am here now' is true in every context. However, the content of 'I exist' at a context can be false at worlds other than the world of the context. For example, the content of 'I exist' in a context C in which Mary is the agent is the proposition that Mary exists; this proposition is true in the world of C, but is false in some other worlds. Therefore, 'Necessarily,

I exist' is false in context C, even though 'I exist' is true in all contexts. Similarly, 'I am here now' is true at every context, but 'Necessarily, I am here now' is false at all contexts whose agent is a thing that could have been at a different location at the time of the context.

Some theorists argue that the above theory of character is not sufficiently fine-grained. Suppose that the content, in a context, of any expression of the form 'that *F*' is the demonstrated object of the context, if that object is *F*. For instance, the content, in a context, of 'that dog' is the demonstrated object of that context, if it is a dog. (This assumption is controversial; see Braun 1994 and King 2001 for discussion.) Now consider the expressions 'that dog' and 'that dog which is black if it is black'. Anything that is a dog is also a dog which is black if it is black. So on the preceding theory, 'that dog' has the same content in every context as 'that dog which is black if it is black'. Therefore, 'that dog' and 'that dog which is black if it is black' have the same character, on the above theory. Yet they seem to differ in meaning in some respect. Some theorists (Braun 1994) propose to use *structured characters* to deal with this apparent problem. The structured character of a complex expression is a structured entity that contains the (functional) characters of the words in the expression as constituents. For instance, the structured character of 'I run' is a structured entity whose constituents include the character of 'I' and the character of 'run'. 'That dog' and 'that dog which is black if it is black' differ in structured character, for the structured character of the latter has the character of 'black' as a constituent, whereas the structured character of the former does not.

References

Arnauld, Antoine. 1662/1964. *The Art of Thinking: Port-Royal Logic*. Translated by James Dickoff and Patricia James. New York: Bobbs-Merrill.

Braun, David. 1994. "Structured Characters and Complex Demonstratives." *Philosophical Studies* 74, 193–219.

Carnap, Rudolf. 1956. *Meaning and Necessity*, second edition. Chicago: University of Chicago Press.

Cresswell, Max. 1985. *Structured Meanings*. Cambridge, MA: MIT Press.

Frege, Gottlob. 1892. "Über Sinn und Bedeutung," *Zeitschrift für Philosophie und Philosophische Kritik*, 100, pp. 25–50. Translated as "On Sense and Reference." In Peter Geach and Max Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. 1952. London: Blackwell.

Hamilton, William. 1860. *Lectures on Logic*. Boston: Gould and Lincoln.

King, Jeffrey. 2001. *Complex Demonstratives: A Quantificational Account*. Cambridge, MA: MIT Press.

Kaplan, David. 1989. "Demonstratives." In Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan*, pp. 481–563. New York: Oxford University Press.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, David. 1972. "General Semantics." In Donald Davidson and Gilbert Harman (eds), *Semantics of Natural Language*, pp. 169–218. Dordrecht: Reidel. Reprinted in David Lewis, *Philosophical Papers Volume 1*, pp. 189–232, 1983. New York: Oxford University Press.

Mill, John Stuart. 1872. *System of Logic*. London: Longman, Green, Reader, and Dyer.

Montague, Richard. 1974. "Pragmatics." In Richmond Thomason (ed.), *Formal Philosophy: Selected Papers of Richard Montague*, pp. 95–118. New Haven, CN: Yale University Press.

Russell, Bertrand. 1996. *Principles of Mathematics*. New York: Norton. Originally published in 1903, London: Allen and Unwin.

Salmon, Nathan. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press.

Soames, Scott. 1987. "Direct Reference, Propositional Attitudes, and Semantic Content." *Philosophical Topics* 15, 47–87. Reprinted in Nathan Salmon and Scott Soames (eds.), *Propositions and Attitudes*, pp. 197–239, Oxford: Oxford University Press.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

1.2

SEMANTICS AND PRAGMATICS

Christopher Gauker

Semantics deals with the literal meaning of sentences. Pragmatics deals with what speakers mean by their utterances of sentences over and above what those sentences literally mean. However, it is not always clear where to draw the line. Natural languages contain many expressions that may be thought of both as contributing to literal meaning and as devices by which speakers signal what they mean. After characterizing the aims of semantics and pragmatics, this chapter will set out the issues concerning such devices and will propose a way of dividing the labor between semantics and pragmatics.

Disagreements about the purview of semantics and pragmatics often concern expressions of which we may say that their interpretation somehow depends on the context in which they are used. Thus:

- The interpretation of a sentence containing a *demonstrative*, as in "This is nice", depends on a contextually determined reference of the demonstrative.
- The interpretation of a *quantified* sentence, such as "Everyone is present", depends on a contextually determined domain of discourse.
- The interpretation of a sentence containing a *gradable adjective*, as in "Dumbo is small", depends on a contextually determined standard (Kennedy 2007).
- The interpretation of a sentence containing an *incomplete predicate*, as in "Tipper is ready", may depend on a contextually determined completion.
- The interpretation of a sentence containing a discourse particle such as "too", as in "Dennis is having dinner in London tonight too", may depend on a contextually determined set of background propositions (Gauker 2008a).
- The interpretation of a sentence employing *metonymy*, such as "The ham sandwich wants his check", depends on a contextually determined relation of reference shifting.
- The interpretation of a *conditional* sentence, such as "If you turn left at the next corner, you will see a blue house at the end of the street", depends on a contextually determined *range of prospects*.

I will call devices such as these the *in-between phenomena*. Insofar as the truth values of sentences involving in-between phenomena depend on the contextually determined values of certain parameters, explication of the in-between phenomena seems to belong to semantics. Insofar as the evaluation of utterances of such sentences depends on the

context of utterance, explication of the in-between phenomena seems to belong to pragmatics.

Toward dividing the labor between semantics and pragmatics, I will urge that we take care to distinguish between two questions. One question concerns the conditions under which a sentence is true relative to a context. That question belongs to semantics. The other question concerns which context *pertains* to a given act of speech. That question belongs to pragmatics. So characterized, the issue concerns not just how to define "semantics" and "pragmatics" but how the truth conditions of sentences and the settings in which sentences are uttered jointly contribute to a successful linguistic exchange.

1 Semantics, Briefly

A semantic theory for a language will specify the literal, conventional meaning of expressions of the language. Typically, it is assumed that the core of such a specification of meanings is a specification of *extensions*. The extension of a proper name might be the thing named. The extension of a verb phrase might be a set of things that the verb phrase describes. The extension of a whole sentence is a truth value. The extension of a quantifier might be a function from sets (sets of sets) to truth values. In simple cases, the extension of a compound expression will be determined by the extensions of its components and its grammatical structure, and, where this is so, a semantic theory is expected to specify the way in which this is determined. Thus, the truth value of "Socrates is snub-nosed" will be *truth* if and only if the extension of "Socrates", namely, Socrates, is a member of the set of objects that is the extension of "is snub-nosed", namely, the set of snub-nosed things.

Beyond identifying the extensions of expressions, a semantic theory might be expected to specify the *intensions* of expressions. Commonly, the intension of an expression is identified with a function from possible worlds to extensions. Moreover, some of the expressions of a language may express operations on intensions. Thus, "possibly" may be thought of as expressing an operation on intensions that takes an intension and generates the value *truth* if and only if that intension is a function from worlds to truth values that generates truth at some world. Propositions are often identified with the intensions of sentences and are thought of as being, or as determining, functions from possible worlds to truth values.

Many of the controversies in semantics concern the treatment of specific linguistic devices within this basic framework. Other controversies have to do with the basic concepts that I have employed in characterizing this framework. For instance, one controversy concerns the use of the notion of possible worlds. Furthermore, one may question the nature of the relation between expressions and their extensions, which we may call the *reference* relation. What makes it the case that "Socrates" refers to Socrates? What makes it the case that "is snub-nosed" describes the objects in the set of snub-nosed things? Some philosophers have attempted naturalistic reductions of the reference relation. Others despair of obtaining a naturalistic reduction and consequently present ideas about how to formulate a semantic theory without our needing a naturalistically reducible reference relation.

In general, a semantic theory will relativize the truth conditions of sentences to evaluations of various parameters, such as interpretation of the nonlogical lexical items, model domain, possible world or world-and-time, or, as we will see, context. Various notions of validity may be defined by picking out a set of these parameters and, while

holding the values of the other parameters fixed, saying that an argument (a set of premises and a conclusion) is *valid* if and only if for each parameter in that set, if the premises are all true relative to that parameter, then so is the conclusion. For instance, we may define one very narrow kind of validity by taking the pertinent set to be the set of all evaluations of all of the parameters (holding fixed only the interpretation of the logical vocabulary). We may obtain a much broader kind of validity by keeping the interpretation, model domain, and world-time fixed and taking the pertinent set to be only the set of all contexts.

Even theorists who disagree about the basic framework for semantics might nonetheless agree that a semantic theory has as one of its aims the characterization of various kinds of validity. That this is so will be important in what follows, because I will appeal to such logical properties in trying to decide between different approaches to the in-between phenomena.

2 Pragmatics, Even More Briefly

Suppose, for the moment, that declarative sentences always literally express complete propositions relative to a context of utterance. Still, what a speaker means *by* uttering a sentence may be more than that the proposition it expresses is true. A speaker may utter the sentence, "There's a gas station around the corner", and if that sentence expresses in context the proposition that there is a gas station around the corner from place *d*, where *d* is the location at which the speaker is speaking, then one thing the speaker may mean by uttering that sentence is that there is a gas station around the corner from *d*. But in addition the speaker may mean that the gas station is open and has gas to sell. Pragmatics may be conceived as, in part, the study of the way in which speakers manage to mean more by uttering sentences than what the sentences they utter literally mean relative to the context in which they are uttered. That, at least, is the conception of pragmatics that we have inherited from Grice (1989).

Inasmuch as meaning something by an utterance of a sentence is a state of a speaker's mind, this way of defining the subject matter of pragmatics assumes that it concerns the speaker's state of mind. A different way of approaching pragmatics would be to think of it as the study of the way in which hearers are able to draw conclusions on the basis of their understanding of the literal meaning of a speaker's words and their own knowledge of the state of the environment in which the speaker's utterance takes place (Gauker 2001). A hearer who hears a speaker say, "There's a gas station around the corner", may infer from the fact that there is a gas station around the corner and the fact that he or she is standing at a busy intersection in the middle of the day that probably the gas station is open and has gas to sell. The speaker may usually intend the hearer to draw such inferences, but it is not obvious that the hearer has to think about what the speaker had in mind in order to do so.

3 The Logic of the In-between Phenomena

Part of the challenge posed by the in-between phenomena is that they introduce logical data that need to be accounted for. We find arguments that are valid *in the sense* that for every context in which the premises are true, the conclusion is true, and other arguments that are invalid *in the sense* that there is a single context in which the premise is true and the conclusion is not. So our account of the truth conditions of sentences must

explain how the truth value of a sentence relative to a context depends on the content of the context in such a way that in terms of that explanation we can explain why some arguments are valid in this sense and others are not.

Imagine a store that sells nothing but unpainted wooden furniture. In conversing with a customer, the salesman gestures toward the showroom and declares, "Everything is made of wood". In the context in which he speaks, we should regard what he says as true. We would not want to deem it not true, as it would be, for instance, if a plastic chair were for sale. And yet, the telephone on his desk is not made of wood. The sentence "The telephone is made of wood" is not true in that context (which is not to say that it is false). So the argument from "Everything is made of wood" to "The telephone is made of wood" does not preserve truth-in-a-context. But if there were any context in which "The telephone is made of wood" were true, then in that context "Something is made of wood" would be true in that context as well.

For another example, in any context in which "Tipper is ready" is true, "Tipper is ready for something" will surely be true. And yet, there might be a context in which "Tipper is ready for something" is true, though "Tipper is ready" is not true. Tipper might be ready to take her next breath but not be ready for the thing that matters in the context, such as going to the convention center.

4 Approaches to the Middle Ground

4.1 Minimalists

One approach to the in-between phenomena is just to say that, while their contribution to the proposition that a sentence expresses is independent of the context of utterance, the propositions expressed are only *minimal* propositions of a certain sort.

One kind of minimalism is that of Emma Borg (2004, 2007). Borg's motivation is the idea that a semantic theory should characterize those aspects of a speaker's knowledge of meaning that can be attributed to a cognitive module that deals with language, while our understanding of contexts cannot be located in any particular cognitive module. Borg will explain the truth condition of a sentence such as "This is red" as follows (2004: 192):

> (R) If a speaker utters a token of "This is red" and refers with the utterance of "this" therein to an object *x*, then that utterance is true if and only if *x* is red.

Borg's theory does not assign to the *sentence* "This is red" a proposition that it expresses, but on her theory one can still say that an *utterance* of "This is red" expresses a proposition, regarding the object referred to by the utterance of "this" therein, that it is red. Borg is prepared to allow that in a sentence such as "Tipper is ready" there is a hidden argument place and a hidden quantification. So she says that "Tipper is ready" has the same "underlying form" as "Tipper is ready for something" (2007: 350–351).

The minimalism of Cappelen and Lepore (2005) is motivated by the desire to draw a line. Cappelen and Lepore allow that the proposition that a sentence expresses may be relative to a context in some ways. A context may be called upon to provide an interpretation for a limited number of items, such as explicit demonstratives, indexicals, words and inflections indicating tense, and a few others (2005: 1). But they expect that the

arguments that might persuade us to countenance varieties of context relativity beyond those on their list would prove too much, leaving us with no account of what we understand in understanding the meanings of words. They hold that a speaker who utters the sentence "There are no French girls" (their example, 2005: 41) might thereby state *that there are no French girls in room 401*, but what the speaker's utterance expresses is precisely the proposition that *there are no French girls*. A speaker who utters "The students are ready" might be interpretable as saying that the students are ready for the exam, but what the speaker's utterance expresses is precisely the proposition *that the students are ready* (2005: 168).

A problem for Borg's theory is that it is not evident that her truth conditions for sentences containing demonstratives can be derived from a recursive definition of truth for a language. For example, the truth condition of "This is *not* red" ought to be a consequence of (R) and a general fact about the truth conditions of negations. But what is this general fact about the truth conditions of negations? We could try:

(N) If *u* is an utterance of a sentence *s* and *v* is an utterance of the negation of *s*, then *v* is true if and only if *u* is not true.

But this would be wrong for two reasons. First, (N) tells us the truth conditions of a negation only on the condition that the sentence negated has been *uttered*, though it is not in general true that if a negation is uttered then the sentence negated has been uttered. And second, we cannot assume, as (N) does, that the demonstratives in the utterance of the sentence negated refer to the same objects as the demonstratives in the utterance of the negation. In place of (N), we might put:

(N′) If *r* is the negation of sentence *s*, then *r* is true in a context *c* if and only if *s* is not true in *c*.

But this belongs to a definition of truth-in-a-context, not a definition of truth simpliciter, whereas it was Borg's ambition to avoid the relativization of truth to context.

Further, neither variety of minimalism accounts for all of the logical facts that the in-between phenomena point to. Neither theory identifies any sense in which the argument from "Everything is made of wood" to "The telephone is made of wood" might fail to be valid. Borg's theory will not explain in what sense "Tipper is ready for something" fails to imply "Tipper is ready" because she holds that these two sentences have precisely the same truth conditions. On Cappelen and Lepore's theory, what we say will depend on what we say about the proposition that *Tipper is ready*, but it seems that for anything that we might say about that, we will end up either affirming that "Tipper is ready for something" implies "Tipper is ready" or denying that "Tipper is ready" implies "Tipper is ready for something", which seem to be wrong results.

4.2 Partialists

A second point of view on the in-between phenomena is that what they show is that sentences frequently do not express complete propositions at all and that semantic theory cannot aspire to anything more than ascribing to sentences a kind of "gappy" meaning that a speaker can utilize to convey complete propositions to a hearer, the *speaker's meaning*. This is the sort of view advanced by Recanati (2001, 2004), Carston (1998, 2008) and Bach (1994, 2005).

The partialists disagree among themselves on several questions. Recanati disagrees with Carston on certain questions concerning the psychology of interpretation (2004: 39). Carston (2008) does not understand why Bach thinks that the interpretation of indexicals, such as "I" and "now", belongs to semantics. And Recanati, in defining a kind of minimal propositional meaning, sometimes takes into account the hearer's point of view (2004: 19); whereas Bach finds no place for such a conception of meaning. Nonetheless, the partialists all agree in thinking that linguistic convention alone will usually not get as far as determining for a sentence, or even for a sentence in a context, a full proposition expressed. What semantics assigns to many sentences is only a *proposition radical* (Bach) or *linguistically encoded meaning* (Carston).

Not only do the partialists not accommodate the special logic of the in-between phenomena, they are altogether silent on questions of logical relations. Inasmuch as sentences do not express propositions, not even relative to contexts, we cannot define logical relations between sentences in terms of relations between the propositions expressed. Carston, for one, holds that only mental representations, not spoken sentences, are the proper object of truth-conditional semantics and that they are free of the in-between phenomena; but to say that is not yet to address the question of the logic of natural language. One can think of various accounts that one might offer on the partialists' behalf. If sentences express some kind of "gappy" meaning, short of a full proposition, then sentence truth could be relativized to ways of filling the gaps, and logical relations could be defined in terms of truth relative to ways of filling the gaps. But, as far as I know, none of the partialists has actually addressed the issue in print.

4.3 Indexicalists

An influential model for the treatment of the in-between phenomena has been Kaplan's treatment of indexicals such as "I" and "now" (1989). According to Kaplan, the truth of a sentence must be relativized to both a possible world and a context. A context, for Kaplan, is an object that assigns to each indexical an appropriate referent. For example, to "now" it assigns a time. Thus, to each expression we can assign a function from contexts to intensions, which in turn may be thought of as functions from worlds to extensions. Kaplan called these functions from contexts to intensions *characters*.

The indexicalist strategy for explicating the in-between phenomena, exhibited in a series of papers by Stanley (e.g., 2002, 2005), is to treat all of them along the same lines as indexicals. The in-between phenomena are to be thought of as involving additional, unspoken lexical items that appear in the deep grammatical structure of a sentence and receive an interpretation from the context. For example, in Stanley and Szabó 2000, domains of discourse for quantifiers are supposed to be sets that contexts assign to variables associated in deep structure with nominal phrases. (I use the term "deep structure" to refer to mental representations of the grammatical structure of a sentence. Contemporary conceptions of this deep structure vary.) Similarly, gradable adjectives, such as "small", could be handled by letting context assign to them a standard, such as an average height (Kennedy 2007). King and Stanley even suggest that the way to handle the context-relativity of conditionals is to let context assign a similarity relation (between possible worlds) to the words "if" and "then" (2005: 152).

The indexicalist strategy suggests a clear division of labor between semantics and pragmatics and also accommodates at least some of the logical facts at issue. A

semantic theory for a language will take the form of a recursively specified assignment of a character, in Kaplan's sense, to each expression (with contexts now extended as necessary). It will fall to pragmatics to explain what it takes for a given context to be the one that *pertains* to a given utterance of a sentence, that is, to be *the* context such that we should evaluate the utterance as true (simpliciter) if and only if the sentence uttered is true relative to that context in the world in which the utterance occurs. The indexicalist can even account for some of the special logical facts due to context-relativity, such as the fact that the argument having "Everything is made of wood" as premise and "The telephone is made of wood" as conclusion is in a sense not valid, by defining the pertinent kind of validity thus: an argument is valid if and only if for each context if the premises are true in that context then the conclusion is true in that same context.

However, it is doubtful whether the indexicalist strategy can explain all of the logical facts pertaining to the in-between phenomena. To see this, consider first a sentence containing two demonstratives, such as:

(1) This is delicious and this is not delicious.

There is nothing contradictory about this sentence, and if it seems odd on the printed page, that is only because we cannot observe the gestures that would tell us what different foodstuffs the two occurrences of "this" denote. To acknowledge that (1) expresses a truth, we do not have to envision that the context somehow shifts between the evaluation of the first conjunct and the evaluation of the second. Sentence (1) may be true in a single context because the two occurrences of "this" may denote different objects in a single context.

By contrast, each of the following sentences is odd in an additional way:

(2) Every student is happy and some student is not happy.
(3) Tipper is ready and Tipper is not ready.
(4) Dumbo is small, and Mickey is large, and Dumbo is larger than Mickey.

There is an oddity in each of these sentences that our semantic theory should enable us to characterize (Gauker 2010). The oddity cannot readily be characterized as merely pragmatic. We can perhaps imagine situations in which one of these sentences would be uttered, but even then the situation is either one in which the speaker has spoken poorly though we can see what he or she has in mind or one in which the oddity is deliberately exploited for special effect (for example, to pose a riddle). The contradictory character of (2)–(4) is a problem for the indexicalist because, for the indexicalist, (2)–(4) will be noncontradictory, just as (1) is, since the context-relative elements in these sentences should be interpretable independently, just as they are in (1).

Incidentally, it has sometimes been supposed that *pure indexicals*, such as "I" and "now", differ from demonstratives such as "this" in that the former but not the latter are governed by strict rules such as that "now" always denotes the time of utterance, and Kaplan himself may have presumed as much. In light of answering-machine cases and post-it note cases (Predelli 2003), that doctrine has been thrown in doubt. Consequently, it is now widely doubted that there is a special logic of indexicals that would ensure that the sentence "I am here now" has a special kind of validity.

5 A Proposal: Contexts Versus Situations

So far I have followed the common practice of using the term "context" equivocally. I have treated contexts as parameters relative to which we define the truth of sentences of a language, and I have spoken of contexts as *contexts of utterance*, which we may think of as situations in which an utterance takes place. The first step toward dividing the labor between semantics and pragmatics in accounting for the in-between phenomena will be to strictly distinguish between contexts proper and situations. A *context* will be a certain formal structure relative to which the truth of sentences will be defined. A *situation* will be an arrangement of objects and events such as may include an *utterance* of a sentence. Semantics will deal with the definition of truth for sentences relative to contexts. Pragmatics will deal with the conditions under which a given context is the one that *pertains* to a given utterance situation.

5.1 Semantics in Light of this Distinction

Although situations are endlessly complex arrangements of objects and events, contexts are precisely definable structures. So we have the possibility of precisely defining the whole set of contexts for a given language. And we have the possibility of defining in a precise way the conditions under which an arbitrary sentence is true relative to an arbitrary context. The logical phenomena pertaining to the in-between phenomena can then be handled by defining an argument to be in a sense valid if and only if it preserves truth-in-a-context and defining a sentence to be in a sense contradictory if and only if it is true in no context.

To accommodate context-relative domains of discourse, we can let a context include a set of objects, the domain (Gauker 1997b). To accommodate the context-relativity of gradable adjectives, we can let contexts include, for each gradable adjective, a standard that an object must surpass in order for the adjective to apply. To accommodate incomplete predicates, we may include a set of propositions and say that, for example, "Tipper is ready" is true in a context if and only if there is some activity such that a proposition to the effect that Tipper is ready for that activity belongs to the contextually-determined set of propositions *and* there is no activity such that the proposition that Tipper is not ready for that activity belongs (Gauker forthcoming). To accommodate metonymy we can let contexts include a deferment function that might, for example, take a ham sandwich as input and yield as output the patron who ordered a ham sandwich. To accommodate the context-relativity of conditionals we might let a context include (or even *be*) a set of prospects (themselves conceived as contexts), so that we can say that an indicative conditional is true in a context if and only if for each prospect in (or identical to) the context, if the antecedent is true in that prospect, then so is the consequent (Gauker 2005).

In light of these characterizations of context and the definitions of logical properties in terms of truth-in-a-context, we can hope to capture the logical data introduced by the in-between phenomena. We can say that the argument from "Everything is made of wood" to "The telephone is made of wood" is invalid, because there is a context *c* such that every object in the domain of discourse for *c* is made of wood but such that the telephone is not in that domain. The sense in which "Tipper is ready for something" fails to imply "Tipper is ready" is that there is a context in which the contextually-determined set of propositions contains the proposition that Tipper is ready for a certain

activity, so that "Tipper is ready for something" is true in that context, but also contains the proposition that Tipper is not ready for a certain other activity, so that "Tipper is ready" is not true in that context.

To account for the contradictory aspect of (2) we may suppose that the context specifies a global domain not assigned to any particular indexical. In (2), the truth of "Every student is happy" in a context demands that the domain for the context include only happy students, while the truth of "Some student is not happy" in context demands that the domain include at least one student who is not happy. (More work has to be done to accommodate the kinds of sentences that motivate Stanley, such as "Every student answered every question" and "Every sailor waved to every sailor". See Gauker 2010.) Insofar as a context contains a single contextually determined set of propositions, we may find that the conditions under which "Tipper is not ready" is true relative to a context are dual to the conditions under which "Tipper is ready" is true relative to a context, so that (3) is true in no context. As for (4), a context may specify a single standard relative to which both "small" and "big" are evaluated, so that if Dumbo is larger than Mickey, then no matter our contextually determined standard for smallness, if Dumbo satisfies it, then Mickey will not satisfy our contextually determined standard for bigness.

5.2 Pragmatics in Light of this Distinction

We may assume that quite generally an utterance of a declarative sentence is true (simpliciter) if and only if the sentence uttered is true relative to the context that pertains to the utterance. In this light, the question for pragmatics becomes: what is it about the situation in which an utterance takes place that determines that a particular context is *the* context that *pertains* to that utterance?

Many authors writing about the in-between phenomena, including all of those mentioned in Section 4, assume that the shortfall between the semantic properties of the sentence uttered and what a given utterance of that sentence expresses is made up by the content of the speaker's communicative intentions. So long as interpretation is conceived simply as filling the gaps left open by literal meaning, it may be hard to conceive of any alternative. But once we conceive of the issue for pragmatics as that of identifying the context that pertains to a given utterance, alternatives become conceivable. One could still insist on the primacy of speakers' intentions, by holding that nothing but the speaker's intention picks out the context pertinent to a speaker's utterance. But an alternative that comes into view is that the pertinent context might be fixed by other features of a situation, such as these: what is perceptually salient in the immediate environment, the facts that are relevant to the interlocutors in light of their goals, what it is true or reasonable to say, and anaphoric connections between context-relative elements in a sentence and prior discourse (Gauker 2008b).

One problem facing the assumption that speaker intention fixes context is that it is just not clear what the contents of the pertinent speaker's intentions are supposed to be. In the case of demonstratives, the answer may be clear enough: the intention that determines that the pertinent context assigns object *o* to "that" is the intention to draw the hearer's attention to object *o*. But it is far less clear what the content of the intention would be that determines the membership of the domain of discourse for the pertinent context. It cannot be, for instance, the intention, for some set *D*, that the sentence uttered be evaluated relative to a context for which the domain of discourse is *D*. That

account of it would be question begging in that it presumes that we have a prior understanding of what it is for a certain set to *be* the domain of discourse for an utterance.

Another basic question for the assumption that speaker intention fixes context is how a hearer is supposed to recognize the speaker's intentions in speaking. Normally, hearers have little access to what a person has in mind apart from an understanding of what the person says. But, as the in-between phenomena indicate, hearers normally have to identify the context pertinent to a speaker's act of speech in order to figure out what the speaker has said. So normally, hearers will have to have some means of identifying the pertinent context without independent access to the speaker's intention. The question then becomes whether speakers have sufficient independent access to the speaker's intention or whether the speaker's intention can determine the pertinent context although the hearers normally identify the pertinent context on some other basis.

Finally, one has to give some thought to the mental representations that are the speaker's underlying intentions. It is not plausible that every aspect of the meaning of an utterance that is not written out explicitly in the sentence uttered is somehow written out explicitly in the brain. The meanings of mental representations will depend on contextually determined referents and contextually determined domains of discourse, much as the meanings of spoken sentences do (Gauker 1997a). So to assume that the speaker's intention is what determines the context pertinent to a speaker's utterance just puts off explaining how something other than speaker's intention can determine the context pertinent to a particular token representation.

6 Beyond the In-between

There are many linguistic phenomena that challenge the core conception of semantics described in Section 1 although they do not appear to belong to the same family as the in-between phenomena listed in the introduction. These include anaphoric pronouns, words that indicate a rhetorical relation between what is being said and prior discourse (such as "but" and "anyway"), and adjectives whose meanings seem to depend on the nouns they modify (such "good" in "good actor"). The proposal of Section 5.1 does not immediately suggest a means of accommodating these phenomena in a compositional, or recursive, semantic theory. But perhaps it gives us reason to hesitate before giving up on the idea that that a compositional or recursive theory of suitable semantic values of some kind, relativized to parameters of some kind, might encompass such phenomena as well.

Related Topics

1.1 Extension, Intension, Character, and Beyond
1.4 Presupposition
1.5 Implicature
1.6 Pragmatic Enrichment and Conversational Implicature
1.10 Context-Sensitivity
1.14 Relevance Theory
3.3 Adjectives
3.6 Anaphora
3.11 Indexicals and Demonstratives
6.1 Philosophy of Language for Epistemology
7.7 Pragmatics and Context: The Development of Intensional Semantics

References

Bach, Kent (1994) "Conversational Impliciture", *Mind and Language* 9: 124–162.
Bach, Kent (2005) "Context *ex Machina*", in Z. Gendler Szabó (ed.) *Semantics vs. Pragmatics*, Oxford: Oxford University Press, pp. 15–44.
Borg, Emma (2004) *Minimal Semantics*, Oxford: Oxford University Press.
Borg, Emma (2007) "Minimalism versus Contextualism in Semantics", in G. Preyer and G. Peter (eds.) *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*, Oxford: Oxford University Press, pp. 339–359.
Cappelen, Herman and Ernie Lepore (2005) *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*, Oxford: Blackwell Publishing.
Carston, Robyn (1998) "The Semantics/Pragmatics Distinction: A View from Relevance Theory", *UCL Working Papers in Linguistics* 10: 1–30.
Carston, Robyn (2008) "Linguistic Communication and the Semantics/Pragmatics Distinction", *Synthese* 165: 321–345.
Gauker, Christopher (1997a) "Domain of Discourse", *Mind* 106: 1–32.
Gauker, Christopher (1997b) "Universal Instantiation: A Study of the Role of Context in Logic", *Erkenntnis* 46: 185–214.
Gauker, Christopher (2001) "Situated Inference versus Conversational Implicature", *Noûs* 35: 163–189.
Gauker, Christopher (2005) *Conditionals in Context*, Cambridge, MA: MIT Press.
Gauker, Christopher (2008a) "Against Accommodation: Heim, van der Sandt, and the Presupposition Projection Problem", in J. Hawthorne (ed.) *Philosophical Perspectives*, 22, *Philosophy of Language*, Oxford: Blackwell Publishing, pp. 171–205.
Gauker, Christopher (2008b) "Zero Tolerance for Pragmatics", *Synthese* 165: 359–371.
Gauker, Christopher (2010a) "Global Domains vs. Hidden Variables", *The Journal of Semantics* 27: 298–328.
Gauker, Christopher (forthcoming) "What Tipper is Ready for: A Semantics for Incomplete Predicates", *Noûs* (forthcoming).
Grice, Paul (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.
Kaplan, David (1989) "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", in J. Almog, J. Perry, and H. Wettstein (eds.) *Themes from Kaplan*, Oxford, Oxford University Press, pp. 481–564.
Kennedy, Christopher (2007) "Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives", *Linguistics and Philosophy* 30: 1–45.
King, Jeffrey C. and Jason Stanley (2005) "Semantics, Pragmatics, Semantic Content", in Z. Gendler Szabó (ed.) *Semantics versus Pragmatics*, Oxford: Oxford University Press, pp. 111–164.
Predelli, Stefano (2003) *Contexts: Meaning, Truth and the Use of Language*, Oxford: Oxford University Press.
Recanati, François (2001) "What is Said", *Synthese* 128: 75–91.
Recanati, François (2004) *Literal Meaning*, Cambridge: Cambridge University Press.
Stanley, Jason (2002) "Making it Articulated", *Mind and Language* 17: 149–168.
Stanley, Jason (2005) "Semantics in Context", In G. Preyer and G. Peter (eds.) *Contextualism in Philosophy*, Oxford: Oxford University Press, pp. 221–253.
Stanley, Jason, and Zoltán Szabó (2000) "On Quantifier Domain Restriction", *Mind and Language* 15: 219–261.

1.3
LOGICAL FORM

Kirk Ludwig

Interest in logical form has been driven historically by interest in valid argument forms, that is, semantic entailment relations between the premises and conclusion of an argument underwritten by their semantic forms. The pressure to provide an ever deeper and more general account of valid argument forms has led to a generalization of the notion of logical form to semantic structure relevant to the evaluation generally of the truth or falsity of a sentence.

It is a staple of the tradition since Frege that logical form differs, sometimes dramatically, from the surface or superficial grammatical form of a sentence. (We return below to whether logical form may be associated with some further level of syntactical analysis.) For example, some sentences, like 2 and 3, have multiple readings ("someone is loved by everyone" and "everyone is such that there is someone he loves"; "relatives who are visiting can be boring" and "it can be boring visiting relatives") which should be assigned different logical forms because the same words contribute differently to truth conditions on each. Some sentences with the same superficial form, such as 1 and 2, 4–6, 9 and 10, 11 and 12, and 13 and 14, are assigned different logical forms because they enter into different systematic entailment relations. And some sentences with superficially different forms, such as 6–8, are assigned the same logical form because they appear to enter into the same systematic entailments.

1. Mary loves John
2. Everyone loves someone
3. Visiting relatives can be boring
4. The hearth is warm
5. The weather is fine
6. The whale is a mammal
7. Whales are mammals
8. Every whale is a mammal
9. Felix does not bark
10. Pegasus does not exist
11. The president is 6' 1" inches tall
12. The average man is 5' 9" inches tall
13. Jack is a wolf
14. Jack ate a wolf

In the early analytic tradition, the divergence of surface and logical form was associated with a divergence between the form of thoughts and the sentences expressing them. Wittgenstein puts this vividly in his *Tractatus Logico-Philosophicus* ([1921] 1961: 4.002).

> Language disguises thought. So much so, that from the outward form of the clothing it is impossible to infer the form of the thought beneath it, because the outward form of the clothing is not designed to reveal the form of the body, but for entirely different purposes.

Bertrand Russell, in the second of his 1914 Lowell lectures, *Our Knowledge of the External World*, defined "philosophical logic" as that portion of logic that concerned the study of forms of propositions—or, as he called them, "logical forms." He claimed that

> some kind of knowledge of logical forms, though with most people it is not explicit, is involved in all understanding of discourse. It is the business of philosophical logic to extract this knowledge from its concrete integuments, and to render it explicit and pure.
>
> ([1914] 1993: 53)

The most famous exemplar of this activity is Russell's Theory of Descriptions (1905), according to which the logical form of "The King of France is bald" is not that of a subject–predicate sentence (or even a noun phrase + verb phrase construction—we return to this divergence later) but a quantificational sentence with internal truth-functional structure, "There is a king of France and whatever is a king of France is identical with it and it is bald," in which the grammatical subject of the original is not treated as contributing any entity to the meaning of the sentence, in contrast to, say, "Louis XIV was bald" (5, 6, and 12 introduce additional complexities).

In the early twentieth century, especially in the short-lived but influential theory of Logical Atomism, logical form was directly connected with metaphysics and the philosophy of mind (Russell [1918–19] 1985; Wittgenstein [1921] 1961). The logical form of sentences was taken to reveal both the forms of possible facts or reality and the thoughts expressed with them. This interest in logical form is clearly broader than an interest in inference patterns among sentences induced by the meanings of logical terms. For Russell, for example, a crucial question about logical form was whether belief sentences express relations to propositions, a view he held at one point but later rejected (Russell [1918–19] 1985: 87–8). (Importantly, Russell's and Wittgenstein's interest in the structure of thought blurs the distinction Davidson drew later between investigation of logical form and conceptual analysis (Davidson [1980] 2001b: 105–6, [1984] 2001c: 31).)

Even apart from Logical Atomism, investigation of logical form is an important component of what Strawson (1959) called "descriptive metaphysics," the project of uncovering the ontology embedded in natural languages, what must exist if the sentences we endorse are true. Given Quine's (1948: 33) criterion of ontological commitment, according to which "a theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true," it follows that one goal of the recovery of logical form is to reveal the quantificational commitments of natural language sentences (see Davidson

1977; Higginbotham 1993). We will consider a dramatic example below in the event analysis of action sentences.

One response to the divergence of surface and logical form is to treat propositions or possible facts as the primary bearers of logical form. Sentences then have the same logical forms if and only if the propositions or possible facts they express do. This gives us, though, only a temporary illusion of progress. As Ayer put it,

> This would be all very well, if we had any means of determining the logical forms . . . other than through the grammatical forms of the sentences which are used to state them. What happens, in practice, is that we decide, on other grounds, which forms of sentences convey this information most perspicuously, and that these are not always the grammatical forms in which the sentences are originally cast.
>
> (1984: 29–30)

Ayer's last remark points to a popular way of talking about logical form, according to which the logical form of a natural language sentence is the form of a sentence in a specially regimented, "ideal," typically formal language, that translates it—or the forms of the sentences that translate it if it is structurally ambiguous (Kalish 1952; Harman 1972; Audi 1995: 511–12; Sainsbury 2001). These sentences are sometimes said *to be*, and sometimes *to give*, the logical forms of the originals. The ideal or regimented language is to have no ambiguities and to encode in its syntax all differences in the logical or semantic roles of terms in them. An interpreted formal language for a first-order logic, designed to make shared validity-relevant semantic features correspond to the construction of a sentence out of its parts, would be an example. Natural language sentences have the same form if and only if they are translated by sentences of the regimented language of the same form.

This is not, however, a satisfactory account, for it leaves us with the question of what the form of the sentence in the ideal language is (Grandy 1974). If we think of it in terms of the pattern of terms in it, perhaps the pattern of logical terms, then this makes the notion of logical form relative to the choice the favored language, and there are many that would do equally well. Russell's analysis of "The King of France is bald" can be rendered into a standard infix logical notation or into Polish or prefix notation (with "f(x)" for "x is a king of France," "b(x)" for "x is bald," and "i(x,y)" in prefix notation for "x = y"), as in 15 and 16.

15. $(\exists x)((f(x) \,\&\, (\forall y)(f(y) \supset x = y)) \,\&\, b(x))$
16. $\Sigma x K K f(x) \Pi y C f(y) i(x,y) b(x)$

The logical form of "the king of France is bald" cannot be identified with either, or with the patterns of the terms in them, since they are distinct and the choice arbitrary. We can only say what the logical form is relative to a system of regimentation. Yet, the original and its translations (if they are) *share* logical form. Sensitivity to this point explains why philosophers often talk of such renderings as *giving* the logical form of the original. But this leaves us with the question of *what* it gives and *how* it is doing it.

The idea that logical form attaches first to propositions points to a resolution, for propositions (in the Frege–Russell tradition) were introduced to be simultaneously the objects of thought and the meanings of (declarative) sentences. This suggests

looking midway between the sentence and the proposition it expresses to a comprehensive semantic theory that details how words and structures contribute to fixing truth conditions for sentences.

An influential suggestion of Donald Davidson's along these lines is that the notion of logical form is best placed in the context of an interpretive truth theory for a natural language.

> What should we ask of an adequate account of the logical form of a sentence? Above all . . . such an account must lead us to see the semantic character of the sentence—its truth or falsity—as owed to how it is composed, by a finite number of applications of some of a finite number of devices that suffice for the language as a whole, out of elements drawn from a finite stock (the vocabulary) that suffices for the language as a whole. To see a sentence in this light is to see it in the light of a theory for its language. A way to provide such a theory is by recursively characterizing a truth predicate, along the lines suggested by Tarski.
>
> (Davidson 1968, [1984] 2001c: 94)

The suggestion in full is that the logical form of a sentence is revealed in the context of a compositional meaning theory for the language that takes the form of an interpretive Tarski-style truth theory (Tarski [1934] 1983, 1944; Davidson 1967b, 1970, 1973). Restricting attention to a context-insensitive language, an interpretive truth theory for an object language L is a finitely axiomatized theory whose axioms provide satisfaction conditions for the semantically primitive expressions of the language by using metalanguage expressions which translate them. With an adequate logic, the theory will entail all sentences of the form (T), where "s" is replaced by a description of a sentence of L as constructed out of its semantically primitive components (the need to regiment natural language sentences introduces a complication which we return to below) and "p" is replaced by a metalanguage sentence that translates it.

(T) s is true_L iff p

In (T), "is true_L" is a metalanguage truth predicate for sentences of L. Given that what replaces "p" translates s, we can replace "is true_L iff" with "means_L that" and preserve truth. A sentence of the form (T) in which this replacement yields a true sentence we call a *T-sentence* and we say that it provides interpretive truth conditions for its object language sentence. (The morals developed extend to a language with context-sensitive elements, though we cannot take the extension up here. See Lepore and Ludwig 2005: chs. 4–5 for more detail and the relation of this way of putting the project to Davidson's own.)

A *canonical proof* issues in a T-sentence for s by drawing intuitively only on the content of the axioms for words in s. The canonical proof of a T-sentence shows the semantic role of each primitive expression in the sentence in fixing its interpretive truth conditions, and in this sense shows what its semantic structure is. According to Davidson's suggestion, this is to know the logical form of the sentence: "To know the logical form of a sentence is to know, in the context of a comprehensive theory, the semantical roles of the significant features of the sentence" ([1980] 2001a: 146).

A simple informal theory, A1–A8, stipulated to be interpretive, will illustrate. The vocabulary of L consists of the variables, "x," "x_1," "x_2," . . ., names "Ned" and "Sally,"

the one-place predicate "is honest," the two-place predicate "loves," the connectives "~" and "&" (for negation and conjunction), and an existential quantifier formed with parentheses around "∃" followed by a variable, e.g., "(∃x)." Square brackets will form a structural description of a sentence. Thus, "[Ned loves Sally]" is equivalent to "'Ned'⌢'loves'⌢'Sally'"—the concatenation of "Ned," "loves," and "Sally." Similarly, where "v" and "v'"are metalinguistic variables for variables and names (terms) of L, "[v loves v']" is equivalent to "v⌢'loves'⌢v'," etc. 'iff' abbreviates "if and only if." We use functions from terms to objects as satisfiers of formulas. Axioms A1–2 specify assignments to names for the functions we quantify over. For any α, [f(α)] refers to the object f assigns to α. "ϕ" and "ψ" take object language formulas as values.

A1. For any function f, f("Ned") = Ned.
A2. For any function f, f("Sally") = Sally.
A3. For any function f, term v, f satisfies$_L$ [v is honest] iff f(v) is honest.
A4. For any function f, terms v, v′, f satisfies$_L$ [v loves v′] iff f(v) loves f(v′).
A5. For any function f, for any formula ϕ, f satisfies$_L$ [~ϕ] iff it is not the case that f satisfies$_L$ ϕ.
A6. For any function f, for any formulas ϕ, ψ, f satisfies$_L$ [ϕ & ψ] iff f satisfies$_L$ ϕ and f satisfies$_L$ ψ.
A7. For any function f, for any formula ϕ, variable v, f satisfies$_L$ [(∃v) ϕ] iff for some f′ such that f′ is a v-variant of f, f′ satisfies$_L$ ϕ.
A8. For any sentence ϕ, ϕ is true$_L$ iff every function f satisfies$_L$ ϕ.

The satisfaction relation generalizes the 'true of' relation. If 'Ned loves Sally' is true, then 'loves' is true of Ned and Sally taken in that order. Thus, a function f that assigns Ned to "x" and Sally to "x_1" satisfies "x loves x_1." A function f′ is a v-variant of f iff f′ is like f except at most in what it assigns to v.

A canonical proof is a sequence of metalanguage sentences such that (i) the last member is a sentence of the form (T) with no semantic vocabulary on the right hand side and (ii) each member of it is an axiom or derived from axioms or previously derived sentences by rules applied so as to permit drawing only on the content of the axioms.

Axioms A1–4 are base axioms; axioms A5–7 are recursive axioms. A1–2 are reference axioms for "Ned" and "Sally." A3–4 are predicate satisfaction axioms, for one- and two-place predicates. A5–6 provide recursive satisfaction conditions for formulas constructed with truth-functional connectives. A7 provides recursive satisfaction conditions for quantified formulas. A8 connects satisfaction with truth of closed sentences. The form of the axioms may be said to give the role of the object language term in determining truth conditions of sentences in which it appears and the logico-semantic form of the vocabulary item for which it is an axiom. Then the forms of the axioms which are used in the proof of a T-sentence for s and how they are employed reveal the logical form of s.

A parameter in characterizing semantic form is the classification of axioms. The framework allows for various classifications. A natural classification for tracing truth-relevant semantic structure is to treat the singular reference axioms as having the same form, axioms for predicates with the same number of argument places as having the same form, and axioms for distinct truth–functional connectives and distinct quantifiers as having the distinct forms.

A proof of a T-sentence for [(∃x)(Ned loves x) & (∃x_1)(x_1 loves Sally)] would instantiate A6 for "&" to this sentence. Then we would instantiate A7 for the existential quantifier to each of the conjuncts on the right hand side. With A4 for "loves" and A1 for "Ned" we can then deduce "For any function f, f satisfies [(∃x)(Ned loves x)] iff some x is such that Ned loves x" and similarly for [(∃x_1)(x_1 loves Sally)] with A4 and A2. Then instantiating A8 to the sentence, a series of valid substitutions yields the T-sentence "[(∃x)(Ned loves x) & (∃x_1)(x_1 loves Sally)] is true iff some x is such that Ned loves x and some y is such that y loves Sally."

What is the point of the exercise? The proof displays the semantic structure of the object language sentence in what axioms of what forms are applied at what points in arriving at a T-sentence for it. The axioms give the type of contribution of the term for which they are axioms, and how they contribute is given by the place in the derivation at which it is applied.

The notion of form we arrive at by this method, as Davidson noted, is relative to both the metalanguage and its logic. We seem to face the same problem as we did earlier in identifying logical form with sentences in a particular ideal language. We want to abstract away from incidental features of canonical proofs. A suggestion extending Davidson's proposal is to fix sameness of logical form of two sentences in their respective languages in terms of their admitting of corresponding canonical proofs (Lepore and Ludwig 2002b: 67).

> A proof P_1 of a T-sentence for s_1 in T_1 corresponds to a proof P_2 for a T-sentence for s_2 in T_2
>
> iff_{df}
>
> (a) P_1 and P_2 are sentence sequences identical in length;
> (b) at each stage of each proof identical rules are used;
> (c) the base axioms employed at each stage are of the same semantic type, and the recursive axioms employed at each stage interpret identically object language terms for which they specify satisfaction conditions (with respect to contributions to truth conditions).

In terms of this notion, the suggestion is that (loc. cit.)

> For any sentences s_1, s_2, languages L_1, L_2, s_1 in L_1 has the same logical form as s_2 in L_2
>
> iff_{df}
>
> there are interpretive truth theories T_1 for L_1 and T_2 for L_2 such that
>
> (a) they share the same logic;
> (b) there is a canonical proof P_1 of the T-sentence for s_1 in T_1;
> (c) there is a canonical proof P_2 of the T-sentence for s_2 in T_2, such that:
> (d) P_1 corresponds to P_2.

This yields an unrelativized characterization of sameness of logical form. It does not tell us what *thing* the logical form of a sentence is but rather when any two sentences in any

languages are the same in logical form. An entity can be introduced using this equivalence relation—the logical form of s_1 in L_1 = the logical form of s_2 in L_2 iff s_1 in L_1 has the same logical form as s_2 in L_2—but this provides no additional insight.

This approach can be extended to imperatives and interrogatives by extending truth-theoretic semantics to fulfillment-theoretic semantics. Sentences have fulfillment conditions of different sorts: truth conditions for declaratives, and compliance conditions for imperatives and interrogatives. Compliance conditions are spelled out recursively using the machinery of the truth theory. This is required even for non-declaratives since they can appear in molecular sentences, and we can quantify into mood markers: e.g., "if you are going to the store, buy some milk," "Invest every penny you earn." Then the definition of a canonical proof can be generalized to a fulfillment theory, and likewise that of a corresponding proof, to generalize the account to nondeclaratives (Lepore and Ludwig 2002b: 74–6; Ludwig 2003).

Our illustration elides an important stage in the application to natural languages. Minimally, structural ambiguity and syntactic elision that leaves inexplicit aspects of how words contribute to truth conditions in natural language sentences require a translation first into a regimented notation to which the truth theory can be applied (Lycan 1984: ch.1). Much of the work of uncovering the logical form of natural language sentences is expressed in appropriate regimentation. In practice, regimentations draw heavily on structures we already know how to incorporate into a truth theory, with the attendant danger that familiar structures will prove a Procrustean bed for the original (we review an example below).

Thus, giving a sentence's logical form with a regimented sentence comes down to producing a sentence whose semantic structure is presumed (a) to be well understood and (b) to be the same as that of the original. The property indicated, however, is exhibited only in the light of the semantic theory for the regimented sentence.

What are the constraints on regimentation? An interpretive truth theory requires that axioms use terms and structures that translate or interpret object language terms and structures. The meaning of a term or a structure is a matter of the rules for its use, realized in speakers' linguistic competencies. These involve both the grammar of sentences and their interpretation, and are expressed in judgments about grammaticality and entailments and in patterns of usage. Recovering these rules from reflection on judgments and observations of usage is constrained (i) by the need to incorporate expressions into a comprehensive semantic and syntactic theory for the language, and (ii) by the need to take into account the role of pragmatic factors in communicative contexts—for the theory must distinguish responses to sentences based solely on meaning and responses based in part on what is pragmatically implied by a speaker (Grice 1989: chs. 1–4). There is no simple general way to describe how to do this, but two examples will give the flavor of the enterprise and show the power of the method.

Return first to Russell's analysis of 'The king of France is bald' in 15, repeated here.

15. $(\exists x)((f(x) \,\&\, (\forall y)(f(y) \supset x = y)) \,\&\, b(x))$

Evidence for this is that the target sentence appears to be true as a matter of meaning if and only if there is a king of France, there is at most one king of France, and whoever is king of France is bald. Additional evidence comes from the power of the analysis to solve certain puzzles. For example, if 17 is a logical truth, an instance of the law of the

excluded middle, how do we avoid the conclusion that it is a logical truth that there is a king of France?

17. The king of France is bald or the king of France is not bald.

The answer on Russell's theory is that the second disjunct has two readings, as 'not' is treated as modifying the predicate (taking narrow scope) or the whole sentence (taking wide scope), as shown in 18a and 18b.

18a. $(\exists x)((f(x) \& (\forall y)(f(y) \supset x = y)) \& \sim b(x))$
18b. $\sim(\exists x)((f(x) \& (\forall y)(f(y) \supset x = y)) \& b(x))$

If we give it the construal in 18a, then 17, though it is committed to there being a king of France, is not a logical truth, while if we give it the construal in 18b, it is a logical truth but not committed to there being a king of France. "The king of France is not bald" is assimilated to "All that glitters is not gold," which has a true and a false reading as we take "not" to modify "gold" or the whole sentence. "The king of France is not bald" then contrasts with "Louis XIV is not bald," where whether "not" modifies the predicate or the whole sentence makes no difference to its evaluation.

Strawson (1950) objected famously that when there is no king of France, an assertion of "The king of France is bald" is not false but lacks a truth value because that there is a king of France is a mere presupposition rather than an entailment of it. For when there is no king of France, we respond to assertions of "The king of France is bald" not by denying it but by saying "There is no king of France." Against this, we feel no hesitation in saying some assertions of sentences containing a nondenoting definite description are false: for example, "My father is the present king of France" (Neale 1990: 26–7). The identified pattern of response, moreover, can be explained on Russell's theory together with a theory of conversational pragmatics (Grice 1989: chs. 1–4). If one denies that the king of France is bald, one will typically be taken to be denying it because one thinks the king of France is hirsute, since the point of the sentence is to pick out something to ascribe a property to it, and, thus, attention is focused on the predicate. If one denies it because there is no king of France or there are two or more, one must be more specific to avoid misunderstanding.

There are other grounds to modify Russell's account, however. Russell offers a construal in terms of unrestricted quantifiers. The practice goes back to Frege, who rendered "All philosophers are rich" as "For all x, if x is a philosopher, x is rich" and "Some philosophers are rich" as "Some x is such that x is a philosopher and x is rich." This introduction of logical connectives can seem unmotivated from the standpoint of syntax. Reflection on related constructions gives substance to this concern and shows the way to a resolution. Prima facie, "All philosophers are rich" and "Most philosophers are rich" should have the same semantic structure, differing only in the quantificational determiner (Neale 1990: 38–44, 2002). However, "Most philosophers are rich" is not logically equivalent to any first-order construal using unrestricted quantifiers (Rescher 1962). It is not equivalent to "Most x are such that if x is a philosopher, then x is rich" because the original is false, while this is true because most things are not philosophers. It is not equivalent to "Most x are such that x is a philosopher and x is rich" because most things are neither philosophers nor rich—nor to any other representation using only unrestricted quantifiers. "Most philosophers" must be treated a distinct semantic unit, "[Most x: x is a philosopher]," which restricts the domain to philosophers and requires most to

satisfy the following predicate. But we should then construe "All philosophers are rich" as "[All x: x is a philosophers](x is rich)" and "Some philosophers are rich" as "[Some x: x is a philosopher](x is rich)" and "The king of France is bald" as "[The x: x is king of France](x is bald)." We can thus retain Russell's quantificational account of definite descriptions without reading any truth functional structure into our understanding of them. We see here an example of how regimentation in a familiar notation can distort logico-semantic form, and how placing the project in the context of a comprehensive semantic theory for the language can provide a correction. For further discussion, see Neale 1990, Reimer and Bezuidenhout 2004 and Ludlow (Chapter 3.7) in this volume.

Another celebrated example of the discovery of logical form is the event analysis of action sentences (Davidson 1967a). The event analysis was introduced to explain the semantic function of adverbs of action, and in particular to explain modifier drop entailment. In 20, the action verb is modified by four adverbials; 20 entails each sentence obtained by removing one or more of them, and together with 21 entails 22.

20. Brutus stabbed Caesar [violently] [with a knife] [at the Curia of Pompey] [on the ides of March]
21. Brutus stabbed Caesar only once.
22. Brutus's stabbing of Caesar was violent, was done with a knife, at the Curia of Pompey, on the ides of March.

That these entailments are independent of the particular adverbials and verb shows that they are a matter of form (this holds of event verbs generally, and of state verbs that admit of adverbial modification). That 20 together with 21 entails 22 is of particular interest because 22 contains a description of an event: namely, the event of Brutus's stabbing Caesar, and "was violent," "was done with a knife," etc., are predicates of the event described. It can scarcely be an accident that variants of these appear in the adverbials in 20. Davidson suggested that these entailments fall into a familiar pattern, which explains in a uniform way the function of the recurring words in 21–22, if we take 20 to involve an implicit existential quantifier over events introduced by the action verb, and the adverbials as contributing predicates of events, as in 23 (we set aside tense for this abbreviated discussion; see Lepore and Ludwig 2002a for a quantificational treatment).

(∃e)(stabbing(e, Brutus, Caesar) & violent(e) & with(e, a knife) & at(e, the Curia of Pompey) & on(e, the ides of March)).

Modifier drop entailment is then an instance of conjunction elimination in the scope of an existential quantifier; 22 is analyzed as in 24, and so "violent," "with a knife," etc., are exhibited as playing the same semantic role in each of these sentences, and the formal entailment of 22 by 20 and 21 is made transparent.

24. [the e: stabbing(e, Brutus, Caesar)](violent(e) & with(e, a knife) & at(e, the Curia of Pompey) & on(e, the ides of March))

In a comment on Davidson's suggestion, Casteñeda suggested separating out the role of agent and object (or patient) of the action as separate conjuncts to accommodate entailments like those between 25 and 26 on the one hand, and 25 and 27 on the other (1967).

25.	He flew the spaceship	(∃e)(agent(e, he) & object(e, the spaceship) & flying(e))
26.	He did something	(∃e)(agent(e, he))
27.	The spaceship flew	(∃e)(object(e, the spaceship) & flying(e))

This has become a standard feature of the event analysis, now the dominate view of the semantics of adverbial modification. (See Ludwig 2010 and Graff Fara and Schein's chapters (2.8 and 3.9) in this volume for further discussion and refinements.)

The event analysis shows vividly the relation between logical form and descriptive metaphysics. If the event analysis is correct, then, as we are committed to the truth of action sentences, we are committed to the existence of events, though there are no overt quantifiers or terms for events in them.

Let us return now to relate this conception of logical form to formally valid arguments. The truth-theoretic (or fulfillment-theoretic) account characterizes logical form in terms of semantic structure. In virtue of the meaning of a sentence with a certain semantic structure (perhaps on a reading), its truth may require the truth of another, as in the case of modifier drop entailment above. Similarly for a set of sentences and one or more sentences. The truth theory does not state these relations. However, the logic of the truth theory, which is needed to carry out proofs of T-sentences, if fully adequate to the meanings of the terms and structures over which it operates, will determine what entailments there are based on the forms of sentences. Thus, the truth theory together with an adequate logic will provide an account of the formal entailment relations among object language sentences.

What is the relation of this to logical validity? An argument is logically valid on the standard account if no uniform reinterpretation of its nonlogical elements, holding fixed the interpretations of its logical terms, makes its premises true without making its conclusion true. Logical validity is therefore relative to a division of terms into logical and nonlogical. However, there is no settled view on the correct extension of "logical constant." While there is general agreement on the requirement that logic be topic neutral, and on the classification of certain terms, there are various ways of developing the idea that lead to different extensions. (See Gomez-Torrente 2002 and MacFarlane 2010 for an overview.) Consequently, there is no settled view on the extension of 'logically valid'.

However, no matter what the criterion, logical validity, if understood in terms of the notion of a logical constant, is narrower than that of formal validity. For example, 28 entails 29. This is intuitively a formally valid argument, but neither 28 nor 29 contain any logical terms.

28. Brutus is an honorable man
29. Brutus is a man
30. Brutus is honorable and Brutus is a man

Of course, the representation of the logical form of 28 in 30 contains a logical constant. But this is to say that 28 and 30 share logical form on the criterion introduced, not that 30 contains any logical terms. On this ground, Evans (1976) distinguished between logical and structural consequences. Logical consequences hinge on the presence of logical terms, structural consequences on the patterns of types of terms in them. Entailments like that from 28 to 29 are structural but not logical. Still, this distinction does not seem

significant from the standpoint of semantic theory. That the pattern is created in the one case by the use of a term and in the other by arrangement of categories of terms seems a matter of what device is used to subserve a purpose. Thus, it seems a matter of terminology whether we speak of logical syntax as a semantic structure characterized indifferently by a pattern formed partly without or entirely without particular terms, or of formally valid arguments, and distinguish them, as Evans does, into those valid in virtue of a pattern created by certain terms, or a pattern created by a structure in the types of expressions used (Lepore and Ludwig 2002b: sec. IV and appendix C).

Returning to logical constants, there seems little point in insisting that one of the competing criteria in the literature identifies the true coin of logic. Each refines the initial observation that there are terms that form salient patterns relevant to validity. When we are clear about what distinctions each draws, nothing further is revealed about the machinery of language by the choice of what terms to bring under the heading "logical." In a comprehensive truth or fulfillment-theoretic account of the language, the role of each expression and structure, whether classified as logical or not, is fully revealed. The choice lies with the purposes for which the distinction is deployed.

A final topic is the relation of nonovert levels of syntactic description to logical form. The first suggestion of an alignment between the theory of syntax and logical form was made by Gilbert Harman (1975), who suggested that deep structure in transformational grammar could be identified with logical form. More recently, with changes in the Chomskian program, it has been suggested that syntactic representations at a level called "Logical Form" or "LF," distinct from surface structure and deep structure, might be identified with the logical forms of sentences (Neale 1993). LF is a level of representation of syntax in Chomskian grammar that provides an interface with a semantics for the language in the sense that the LF representation makes explicit whatever is relevant to semantic interpretation, such as quantificational structure (May 1985, 1987, 1989). This motivates calling it "Logical Form."

That there are levels of syntactic representation, how many, and how realized, is an empirical issue (many linguists, following Chomsky (1995), now posit only Phonetic Form and LF). As usually understood, a level of syntactic representation, distinct from surface form, is a psychologically real level of representation by speakers over which rules of syntax are defined. It is conceivable that the hypothesis of LF be shown to be empirically inadequate. This would not, however, be a reason to say that sentences did not have logical form. The notion of the logical form of a sentence, then, is not to be analyzed in terms of LF. We can also note that our earlier argument against identifying logical form with a sentence in a formal language which makes explicit its semantic structure, relative to a semantic theory, applies here as well. Still, LF is clearly relevant to logical form. If LF is real, then a psychologically real level of syntactical description encodes structural semantic features of sentences. This provides a representation that can guide regimentations that serve as input to an interpretive truth theory and an important empirical constraint on them (see e.g. Hornstein 2002). Work on LF, then, if LF has the relevant properties, will interact in a straightforward way with a theory of logical form for natural languages, and vice versa.

Related Topics

1.2 Semantics and Pragmatics
1.8 Compositionality

1.10 Context-Sensitivity
2.2 Theories of Truth
2.8 Event Semantics
3.4 Quantifiers and Determiners
3.7 Descriptions
3.9 Adverbs
3.11 Indexicals and Demonstratives
4.3 The Role of Psychology
4.6 The Role of Intuitions
4.2 The Role of Linguistics
6.3 The Philosophy of Language for Metaphysics
7.4 Frege, Russell, and Wittgenstein.

References

Audi, R. (ed.) (1995) *The Cambridge Dictionary of Philosophy*, Cambridge: Cambridge University Press.
Ayer, A. J. (1984) *Philosophy in the Twentieth Century*, New York: Vintage Books.
Casteñeda, H.-N. (1967) "Comments on Donald Davidson's 'The Logical Form of Action Sentences'," in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press.
Chomsky, N. (1995) *The Minimalist Program*, Cambridge: MIT Press.
Davidson, D. (1967a) "The Logical Form of Action Sentences," in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh. Reprinted in (Davidson 2001b).
——. (1967b) "Truth and Meaning," *Synthese*, 17: 304–23. Reprinted in (Davidson 2001c).
——. (1968) "On Saying That," *Synthese*, 19: 130–46. Reprinted in (Davidson 2001c).
——. (1970) "Semantics for Natural Languages," in *Linguaggi nella Societa e nella Tecnica*, Milan: Comunita. Reprinted in (Davidson 2001c).
——. (1973) "In Defense of Convention T," in *Truth, Syntax and Modalit*, Amsterdam: North-Holland Publishing Company. Reprinted in (Davidson 2001c).
——. (1977) "The Method of Truth in Metaphysics," *Midwest Studies in Philosophy*, 2: 244–54. Reprinted in (Davidson 2001c).
——. (2001a) "Comments and Replies," in *Essays on Actions and Events*, New York: Clarendon Press. Original publication, 1980.
——. (2001b) *Essays on Actions and Events*, 2nd ed., Oxford: Clarendon. Original publication, 1980.
——. (2001c) *Inquiries into Truth and Interpretation*, 2nd ed., New York: Clarendon Press. Original publication, 1984.
Evans, G. (1976) "Semantic Structure and Logical form," in J. McDowell and G. Evans (eds.) *Truth and Meaning*, London: Oxford University Press.
Gomez-Torrente, M. (2002) "The Problem of Logical Constants," *The Bulletin of Symbolic Logic*, 8 (1): 1–37.
Grandy, R. (1974) "Some Remarks about Logical Form," *Noûs*, 8 (2): 157–64.
Grice, P. (1989) *Studies in the Way of Words*, Cambridge: Harvard University Press.
Harman, G. (1972) "Logical Form," *Foundations of Language*, 9: 38–65.
——. (1975) "Deep Structure as Logical Form," in D. Davidson and G. Harman (eds.) *Semantics of Natural Language*, Dordretch: Reidel.
Hornstein, N. (2002) "A Grammatical Argument for a Neo-Davidsonian Semantics In Logical Form and Language," in G. Preyer and G. Peter (eds.) *Logical Form and Language*, Oxford: Clarendon, 345–64.
Higginbotham, J. (1993) "Grammatical Form and Logical Form," in J. Tomberlin (ed.) *Philosophical Perspectives: Language and Logic*, Atascadero: Ridgeview.
Kalish, D. (1952) "Logical Form," *Mind*, 61: 57–71.
Lepore, E., and K. Ludwig (2002a) "Outline of a Truth Conditional Semantics for Tense," in Q. Smith (ed.) *Tense, Time and Reference*, Cambridge: MIT Press.
——. (2002b) "What Is Logical Form?," in G. Preyer (ed.) *Logical Form and Language*, Oxford: Clarendon Press.
——. (2005) *Donald Davidson: Meaning, Truth, Language, and Reality*, Oxford: Oxford University Press.

Ludwig, K. (2003) "The Truth about Moods," in G. Preyer, G. Peter and M. Ulkan (eds.) *Concepts of Meaning: Framing an Integrated Theory of Linguistic Behavior*, Dordrecht: Kluwer Academic Publishers.
——. (2010) "Adverbs of Action," in T. O'Connor and C. Sandis (eds.) *Blackwell Companion to the Philosophy of Action*, Oxford: Wiley-Blackwell.
Lycan, W. (1984) *Logical Form in Natural Language*, Cambridge: MIT Press.
MacFarlane, J. (2010) "Logical Constants," in E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*
May, R. (1985) *Logical Form: Its Structure and Derivation*, Cambridge: MIT Press.
——. (1987) "Logical form as a Level of Linguistic Representation," in E. LePore (ed.) *New Directions in Semantics*, London: Academic Press.
——. (1989) "Interpreting Logical Form," *Linguistics and Philosophy*, 12: 387–435.
Neale, S. (1990) *Descriptions*, Cambridge: MIT Press.
——. (1993) "Logical Form and LF," in C. Otero (ed.) *Noam Chomsky: Critical Assessments*, London: Routledge.
——. (2002) "Abbreviation, Scope, Ontology," in G. Preyer and G. Peter (eds.) *Logical Form and Language*, Oxford: Clarendon.
Quine, W. V. O. (1948) "On What There Is," *Review of Metaphysics*, 2: 21–38.
Reimer, M., and A. Bezuidenhout. (2004) *Descriptions and Beyond*, Oxford: Oxford University Press.
Rescher, N. (1962) "Plurality Quantification," *Journal of Symbolic Logic*, 27: 373–4
Russell, B. (1905) "On Denoting," *Mind*, 14: 479–93.
——. (1985) *The Philosophy of Logical Atomism*, La Salle: Open Court. Originally published 1918–19 in four consecutive issues of *The Monist*, from a transcript of lectures delived Jan.–Mar. 1918.
——. (1993) *Our Knowledge of the External World: As a Field for Scientific Method in Philosophy*, New York: Routledge. Original publication, 1914.
Sainsbury, M. (2001) *Logical Forms*, 2nd ed, Oxford: Blackwell.
Strawson, P. F. (1950) "On Referring," *Mind*, 59: 320–44.
——. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.
Tarski, A. (1944) "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research*, 4: 341–76.
——. (1983) "The Concept of Truth in Formalized Languages," in *Logic, Semantics, Metamathematics*, Indianapolis: Hackett Publishing Company. Original publication, 1934.
Wittgenstein, L. (1961) *Tractatus Logico-Philosophicus*. Edited by D. F. Pears and B. F. McGuinness, London: Routledge & Kegan Paul Ltd. Original publication, 1921.

1.4
PRESUPPOSITION

Paul Dekker

1 Introduction

A presupposition is normally a proposition that is or has to be taken for granted, or 'supposed', before ('pre-') a certain statement or other linguistic act can be made sense of. A presupposition is not so much the ground you stand on, or the thinking that you rely on, but more like what comes before the ground you stand on, or what comes before the thinking that you rely upon. The notion of presupposition is therefore in essence quite a philosophical subject. By the same token, presuppositions are things which one can try to communicate, without actually stating them. "Has Alfred stopped smoking pot?" Even if you do not know the answer to a question like this, it can be used to communicate, not question, that Alfred used to smoke pot.

Notwithstanding, the term has gotten its main import as a decent technical term, in philosophy and linguistics, and that is how we will mainly deal with it in this chapter. We will not go into the rhetorical uses, and abuses, of the concept. In this introductory section we will speak of the prehistory of the subject and in the remainder we follow, roughly, the historical development. In section 2, the original logico-philosophical interests in and rejections of presupposition are discussed. In section 3, we give a short overview of the major systematic treatments of presupposition in logic and linguistics in the 1970s. In section 4, we focus on more pragmatic and dynamic theoretical treatments of presuppositions that evolved from that, and on some of the most debated issues in this area. In section 5 we very briefly discuss some recent developments.

While the subject of presupposition is as old as the phenomenon of speaking and thinking, it has not become known as such until rather recently. Some idea of presupposition has been commonly attributed to the philosopher Aristotle, although it seems he has never identified it as such. One of the standard (Aristotelian) logical insights is that both of the following statements:

(1) All eudeontologists are herbivors.
(2) No eudeontologists are herbivors.

carry the entailment that there are eudeontologists. And so do the following statements:

(3) Not all eudeontologists are herbivors.
(4) Some eudeontologists are herbivors.

While the examples (1) and (3) seem to be contradictory, as do the examples (2) and (4), they all four seem to entail that there are eudeontologists. If there were no

eudeontologists, all four examples would be deemed unfortunate. The conclusion then seems to be that a sentence can fail to be true, and that also its negation fails to be true, so that there is a gap which has to be bridged for the sentence to be true or false. The gap that has to be filled is the sentence's presupposition. Only if the sentence's presuppositions are satisfied, will we be able to say that the sentence is true or false.

Besides the above-mentioned existential presupposition, to Aristotle (384–322 BC) may be attributed the concept of categorial presupposition as well. Certain predicates cannot be predicated of all kinds of substances, because they are of the wrong type (Aristotle 1975, 2b8–2b22). Nevertheless, the concept of presupposition was not recognized as such by Aristotle. He has been rather explicit that there is this principle of the excluded third, which means that a sentence is either true or false, or at least that its affirmation or its negation must be true. (Aristotle 1975, 17a26ff.)

It has been said that Aristotelian logic has remained unchallenged for 2.5 millennia, and that the concept of presupposition had not been identified over the very same period in the Western tradition. Some possible and notable exceptions are philosophers and theologicians like Eubulides, Thomas Aquinas, and Petrus Hispanus. Eubulides (fourth century BC) was one of Aristotle's contemporaries and antagonists. He proposed the subject of presupposition as a challenge to Aristostelian logic in a form which has become famous as the horned man. "Another was, if you have not lost a thing, you have it; but you have not lost horns; therefore, you have horns" (Laertius 1853, Book VII, chapter "Life of Chrysippus", section XI); see also Laertius 1853, Book II, chapter "Life of Euclides", section IV). This reads like a prehistorical practical joke, but it was a severe problem for Aristotelian logic. If we, as Aristotle seemed to do, endorse the principle of bivalence, at least one of the following sentences must be true:

(5) You have lost your horns.
(6) You have not lost your horns.

Both sentences, however, seem to entail that you have or had horns. If the principle of bivalence holds, one of the sentences must be true, and, therefore, logic tells us that you have or had horns. Since we could have embedded any kind of presupposition beside you having horns, this means that logic would entail any proposition, a pretty disastrous result. (Seuren 2005, following Kneale & Kneale 1962, claims that, with six other challenges, Eubulides has defined the agenda of twentieth century semantics.)

Thomas Aquinas, a thirteenth century theologist and philosopher, apparently adopted a metaphysical or epistemological concept of presupposition, though clearly not a linguistic one: "As a result, every science has presuppositions which the learner must believe" (Aquinas 2000–2009, Faith, Reason, and Theology, Questions I–IV of the Commentary on Boethius' De Trinitate) "For in every science, the nature of its object is presupposed" (Aquinas 2000–2009, "Summa Theologiae, Treatise on Sacred Doctrine"). Larry Horn (Horn 1985; Horn 1996), and many others following him, have credited Pedro Julião (a.k.a. Petrus Hispanus, or Pope John XXI, also thirteenth century) with the acknowledgment of some notion of presupposition as opposed to a notion of denotation. However, Pieter Seuren (Seuren 2005) has pointed out that not only the reference to Peter of Spain is mistaken but also that there is no mention in the text of a notion of presupposition as we understand it nowadays.

2 Logico-Philosophical Treatments

The concept of presupposition as we understand it nowadays was first identified, formally, by the end of the nineteenth century, in the work of Gottlob Frege (1848–1925), in an analytical, and rather positivist, frame of mind. It appears he first spotted the problem in his Foundations of Arithmetic. "Der Ausdruck 'der grösste achte Bruch' hat z. B. keinen Inhalt, weil der bestimmte Artikel den Anspruch erhebt, auf einen bestimmten Gegenstand hinzuweisen" (Frege 1884, p.87–8). ("The expression 'the largest real fraction' has, for instance, no content, because the definite article raises the claim to refer to a definite object.") (Note, the translations from German are mine, PD.) In his groundbreaking "Über Sinn und Bedeutung" Frege literally employed the terms "Voraussetzung", [i.e., presupposition]: ". . . wir setzen eine Bedeutung voraus. . . . Nun können wir uns in jener Voraussetzung freilich irren, und solche Irrtümer sind auch vorgekommen" (Frege 1892, pp. 31–2) (We presuppose a reference. . . . We can easily be mistaken about such presuppositions, and such mistakes have occurred.) "Wenn man also behauptet, 'Kepler starb im Elend', so ist dabei vorausgesetzt, daß der Name "Kepler" etwas bezeichne; aber darum ist doch im Sinne des Satzes "Kepler starb im Elend" der Gedanke, daß der Name "Kepler" etwas bezeichne, nicht enthalten. (. . .) Daß der Name "Kepler" etwas bezeichne, ist vielmehr Voraussetzung ebenso für die Behauptung "Kepler starb im Elend" wie für die entgegengesetzte." (Frege 1892, p. 40) (If one asserts, "Kepler died in misery," it is presupposed that the name "Kepler" denotes something; but the thought that the name "Kepler" denotes something, is therefore not contained in the sense of the sentence "Kepler died in misery". . . . It is more a presupposition for the assertion "Kepler died in misery" that the name "Kepler" denotes something, as likewise for its opposite.) With the logicistic ideal, Frege was not particularly fond of the possibility of presupposition failure, and called it a "Mangel", a deficiency, of natural language: "Dies liegt also an einer Unvollkommenheit der Sprache, von der übrigens auch die Zeichensprache der Analysis nicht ganz frei ist (. . .). (Frege 1892, p. 41) (This is to blame a deficiency of language, from which also the analytic language is not free.) In mathematics, logic, and science, propositions or sentences ought to be true or false and presupposition failure prohibits them from being so. In this, Frege was both idealistic and realistic about the inherent deficiencies of natural language. In his legacy, in a note titled "Meine grundlegenden logischen Einsichten", dated 1915, we read: "Die logische Arbeit ist gerade zu einem grossen Teil ein Kampf mit den logischen Mängeln der Sprache, die uns doch wieder ein unentbehrliches Werkzeug ist." (Frege 1969, p. 272) (Most of the logical work consists in a struggle with the logical defects of language, which, again, is an indispensable tool.)

Bertrand Russell, and Ludwig Wittgenstein, also seem to have realistically acknowledged the problem of presupposition, but idealistically sought to exclude it, either in a reformation of natural language (Russell), or even in the logical analysis of language itself (Wittgenstein). Bertrand Russell (1872–1970) is famous, for his theory of descriptions (1905), which seems to have been the dominant theory until Peter Strawson (on whom more in section 4) attacked it in 1950. A couple of years later, in 1957, Russell, quite irritated, replied:

> "This brings me to a fundamental divergence between myself and many philosophers with whom Mr. Strawson appears to be in general agreement. They are persuaded that common speech is good enough not only for daily life, but also for philosophy. (. . .) He [Strawson] admits that the sentence ["the King

> of France is wise"] is significant and not true, but not that it is false. This is a mere question of verbal convenience. He considers that the word "false" has an unalterable meaning which it would be sinful to regard as adjustable, though he prudently avoids telling us what this meaning is. For my part, I find it more convenient to define the word "false" so that every significant sentence is either true or false." (Russell 1957, pp. 387–9)

Despite these logical and logical positivist reservations, logicians like Jan Łukasiewicz (in 1920), Dmitri A. Bochvar (in 1937), and Stephen Cole Kleene (in 1938, 1952), among many others, have studied many-valued logics, which explicitly recognize the systematic possibility that a sentence be neither true nor false, or be assigned another value than those two. Bochvar proposed the third value to be understood as "uninterpretable," or "nonsense." The consequence of this being that if part of a sentence is rendered nonsense, then the interpretation of the whole crashes and it is rendered nonsense as well. Bochvar's interpretation has a very intuitive understanding, and it has interesting mathematical and computational applications, but the basic idea, understood as an account of presupposition (which it was not intended to be), falls short when we start considering more involved, quantified, or modal structures. It seems to generate the prediction that an existentially quantified sentence will fail to be true or false if the sentence is uninterpretable under one of the assignments to the variable quantified over. Thus, "Someone's bike is stolen" would not need to be true if someone's bike is stolen, but it would be rendered uninterpretable if someone 'failed' a bike.

Bochvar's system has also become known as a weak Kleene logic, in contradistinction to a strong Kleene logic, presented by Kleene himself. In Kleene's (strong) system, the third value is interpreted as "undecided" or "to be determined" and this makes a conceptually different interpretation of the connectives. If, in this system the values of one subordinate formula are not yet determined, we may still be able to assign a determinate value to the whole. So, for instance, if one conjunct of a conjunction is known to be false, we can establish that the whole conjunction will come out false, no matter what the other conjunct has as a value. Likewise, if one member of a disjunction is true, the whole can be known to be true as well. Obviously, a strong Kleene logic, like a weak Kleene one, agrees with a classical propositional logic, as long as the embedded formulas are all defined, or decided. But it also allows for determinate truth values for formulas, one of whose constituent formulas is not (yet) determined. Conceptually, and also computationally, this is an attractive feature. However, it is not obvious how to generalize this intuition when we start considering more involved, quantified or modal structures. It seems to generate the prediction that an existentially quantified sentence may come out not to be false, even if it is undefined for many, if not for all, of the assignments to the variable quantified over. Thus, "Someone's bike is stolen" would not be false if none of those who are 'known' to have a bike, have their bike stolen.

Sure enough, multi-valued approaches need not be implemented and generalized as suggested above. Recent work, by, for instance, Danny Fox (Fox 2008), builds in a sophisticated and empirically very interesting method on that tradition. Some of the work goes back to (Soames 1989), which convincingly shows that presuppositions are not symmetrically projected in coordinations. The presuppositions of a second conjunct or disjunct are not automatically those of a whole conjunction or disjunction. Asymmetric truth tables to that effect have already been proposed by Stanley Peters (Peters 1977). The presuppositions of a second conjunct may get satisfied by those of the first,

in an asymmetric way; the presuppositions of a conditional consequent clause may get satisfied by the material in the antecedent clause of the conditional; and a second disjunct presupposition may get satisfied by the denial or negation of the first. As a matter of fact, Peters' and Soames's truth tables predate some of the dynamic semantic insights which will be discussed in section 4.

3 The Linguistic Decade

Before such a dynamic or pragmatic turn could take over, however, the phenomenon had to be studied in much more detail, and much more extensively. Here is a little piece of discourse, filled with a couple of standard presupposition triggers, most of them identified in the 1970s.

(7) Pete found the lost pack of cigarettes. He managed to hide it from his parents. Ron realized what a good boy Pete was. All the kids thought they were alone. HARRY smoked. It was Bert who lit the cigarette. Mary smoked, too. No, only boys smoked. Amelia had stopped smoking.

The first sentence presupposes that there was some special pack of cigarettes. The second presupposes that there is a referent for the pronoun "it," apparently the pack of cigarettes; the pronoun "he", apparently refers to Peter (who has parents) and who first tried, and then succeeded in hiding the cigarettes from them. The third sentence (counterfactually) presupposes that Pete was a good boy. The fourth presupposes that there was a definite group of kids being talked about. The fifth presupposes that at least someone smoked. The sixth is that someone lit the fatal cigarette. The seventh presupposes that somebody else beside Mary smoked, and the eighth that boys smoked. The last sentence presupposes that Amelia has been smoking earlier.

The presuppositions above are triggered in quite a different number of ways, and people disagree as to whether we can speak of a unitary phenomenon. See (Seuren 1985; Zeevat 1992) for critical discussion. Another problem, however, has been how triggered presuppositions project from embedded positions. A ground-breaking paper on this was produced by Lauri Karttunen (Karttunen 1973), and an excellent overview of this problem is given in (Soames 1989). Karttunen introduced the terminology of 'plugs', 'holes', and 'filters'. So-called holes are sentential operators which inherit the presuppositions of the sentences they are applied to. Thus, all of the following constructions seem to come with the presupposition that Alfred has a niece:

(8) Alfred didn't meet his niece.
(9) Maybe he managed to make an appointment with his niece.
(10) But if he made an appointment with his niece, she didn't show up.

The general idea is that all three sentences entail that Alfred has a niece, even though it is not asserted as such. The conclusion is that "not," "maybe," and "if" are holes, which let presuppositions 'speak up'. This is quite different from so-called plugs which are claimed to literally block presuppositions.

(11) Alfred said he did meet his niece.
(12) He really believes he ate ice with his niece.

Even though these sentences strongly suggest that Alfred has a niece, one cannot be blamed for speaking falsely if Alfred fails a niece. This means that a presupposition of a sentence *S* need not be a presupposition of a sentence *say that S*, or *believe that S*, and it has therefore been claimed that 'saying that' and 'believing that' are presupposition plugs. Presupposition 'filters' are the most difficult and disputed terms. *If . . .*-clauses are typical examples of filters, which sometimes do and sometimes do not inherit the presuppositions of the consequent clauses. Compare the following two sentences.

(13) If Ben marries Lisa, his sons will be in good hands.
(14) If Ben has children with Lisa, his sons will be in good hands.

Most people agree that the first, but not the second, presupposes that Ben has sons. The conclusion seems to be that an *if*-clause may help to filter out a presupposition of a superordinate clause. The data are very much disputed, as will probably appear from the next section.

As a next step to the next section, we may have to note that linguists in the 1970s however already agreed that presupposition is not a purely logical phenomenon. Ruth Kempson (Kempson 1975) argued that presupposition is basically a pragmatic phenomenon, and the conclusions of the leading theorists like Gerald Gazdar, Lauri Karttunen, and Stanley Peters are equivocal. These authors treat presupposition in conjunction with conventional implicatures, which cannot be done without reference to a theory of pragmatics. See, for example, Gazdar 1979. Karttunen and Peters eventually presented an account of presupposition by separating presupposition (or 'implicature') and assertion ('extension') along different dimensions (Karttunen & Peters 1979). See Dekker 2008 for a recent dynamic implementation of their views.

4 The Dynamic Semantic Turn

Taking some steps back in history, it may said to have been Peter F. Strawson (1919–2006) who first formally, identified the concept of presupposition under this name, after Frege did so under the name of "Voraussetzung" (Strawson 1952, p. 175). Before this, in 1950, Strawson had supplied his famous response to Russell's theory of descriptions, noting that it is not sentences that are true or false and terms that refer but that it is the uses of sentences which are true or false, or neither, and people's uses of terms that refer, or fail to refer—and uses which may yield truth-value gaps. (Strawson 1950, pp. 327–8) Above we have already seen Russell's touchy reaction to these insights.

Strawson's emphasis on the use-based nature of truth, reference, and presupposition has dominated large parts of the philosophical/linguistic literature ever since. Strawson's core ideas became more fashionable in linguistics following the work of Keith Donnellan (Donnellan 1966; Donnellan 1978) who once more emphasized the referential uses, as opposed to their Russellian, attributive uses, of definite descriptions. In his seminal paper on presupposition and assertion, Robert Stalnaker continued this line of thought in the early 1970s. Stalnaker emphasizes that "People, rather than sentences or propositions are said to have, or make, presuppositions in this [pragmatic] sense. . . . To presuppose a proposition in the pragmatic sense is to take its truth for granted, and to assume that others involved in the context do the same" (Stalnaker 1970, p. 279). The understanding of presuppositions has gained more emphasis in (Stalnaker 1978), where the notion has been more or less formally implemented in terms of updates of common

grounds, or in terms of a dynamic notion of meaning. In the latter paper, assertions are seen to be, basically, proposals to add propositions to a propositional ground assumed to be common to all participants in a conversation.

The 1980s and 1990s have witnessed several attempts to formalize these pragmatic conceptions of presupposition further but in different, and rival directions, all of them with their own intuitively appealing inclinations. Lauri Karttunen (Karttunen 1974), Irene Heim (Heim 1983), David Beaver (Beaver 1995), and Robert van Rooij (van Rooij 2005), among many others have elaborated and refined analyses based on the idea that presupposition relates to information presumed to be given in a discourse situation. The idea is very simple and intuitively appealing. Consider the following examples.

(15) Julio's son will be a star.
(16) Julio has a son, and Julio's son will be a star.
(17) If Julio has a son, Julio's son will be a star.
(18) Julio's son will be a star, and Julio has a son.

Example (15), but not example (16), presupposes that Julio has a son; and ecample (17) does not, and (18) does, presuppose that Julio has a son. A dynamic semantic treatment of 5 assertion and presupposition neatly accounts for these facts, because such treatments build on the assumption that when the second constituent of a coordinated (conjoined, or implicational) structure is evaluated, including its presuppositions, this is against the background of what the first constituent has implied. And not the other way around. This intuitively explains the asymmetry in the examples above. The main tenets of such dynamic treatments consist in the idea that participants in a conversation maintain some kind of a common ground of what they tend to be commonly given as information. This gets updated any time a contribution to the discourse has been made and has remained unchallenged. Notice that this notion of a 'common ground' is a technical point, and can be suppositional, for instance, if the consequent clause of a conditional sentence is evaluated relative to a suppositional update of the common ground with the antecedent clause.

Within the framework of *Discourse Representation Theory* (*DRT*) a similar, dynamic treatment of presupposition and of presupposition projection has been given by Rob van der Sandt, Bart Geurts, and Hans Kamp (van der Sandt 1988; van der Sandt 1992; Geurts 1999; Kamp 2001). In these, and similar frameworks, the current states of discourses and dialogues are presented as representations of the information conveyed at the current state of a discourse. As in the dynamic theories of interpretation, presuppositions are supposed to be satisfied, but they are not directly identified as propositions, but, again, as representations of them. This enables them to be handled more as independent structures. Within *DRT*, presuppositions are not interpreted as such, but they are handled as representational items that need to be 'resolved' in the context. Consider a complicated structure, like the one induced by the following sentence:

(19) Lucy thinks it is not likely that if she fails her exam, she will not be able to feed her cat.

The crucial, and triggering, part here is "feed her cat" which comes with the presupposition that some y is Lucy's cat. Maybe we all know Lucy has a cat, or maybe we all know Lucy thinks so, or maybe we all know that if what Lucy believes is true, and she also fails

her exam, she has a cat. Or, even, maybe we all know, that Lucy thinks that if she fails her exam than most likely she will not have a cat and be able to feed it. These are all interpretative options made available in *DRT*, surviving from the presupposition that some *y* is Lucy's cat, on a most suitable place in the *Discourse Representation Structure*, where it renders "Lucy's cat" as a known entity right there where the term has to be interpreted. In *DRT* the different resolutions are achieved by moving the representation of *y* being a cat of Lucy's, around in the *Discourse Representation Structure* until it ends in the rationally most comfortable place.

The difference between information being *given* in an update semantics, or being *represented* in a representational framework, like that of *DRT*, may seem marginal, but it has given rise to harsh disputes about the actual implementation of these ideas. A most typical example of the difference is the phenomenon of conditional presuppositions. The strength and beauty of the update approach is that it comes with its own logic, which, almost automatically it seems, explains the facts about presupposition projection. It has been argued, however, that the predictions are too soft to be realistic. Using the dynamic approach, an assertion of one of the following sentences

(20) Lucy is taking a shower now, and she forgot to feed her cats.
(21) If Lucy is taking a shower now, she forgot to feed her cats.

is not predicted to presuppose that Lucy has cats, but only that if she is taking a shower now, she has cats. The predictions are rather counterintuitive. It seems the representational format of *DRT* is quite capable of handling the problem. Upon their most favorable readings both sentences get resolved by unconditionally getting the information that Lucy has cats by resolution or by accommodation. Of course, several good arguments have been made to rescue the dynamic approach. Geurts 1996 gives a critical overview of the problems, and van Rooij 2007 a constructive reply. Despite the apparent contradictory predictions of the two frameworks, which might be overcome, it seems that only (Zeevat 1992) has been able to avoid both horns of the dilemma, and combine the benefits of both types of dynamic approach.

One problem, which the dynamic semantic approach shares with the previously mentioned, many valued approach. consists in the treatment of presuppositions in modal and quantified structures. A most straightforward implemenation of the dynamic ideas seems to entail that, again,

(22) A fat man was pushing his bicycle.

presupposes that all fat men do actually have a bicycle, an unwanted result for sure. Several attempts have been made within *DRT*, and dynamic semantic alternatives, to deal with these problems, but no consensus has been reached. It seems that the intuitive and empirical data available do not point at a direction for solving the issue.

5 Recent Developments

Philosophers and linguists have equally impressively emphasized the phenomenon of presupposition as a pragmatic one, but it has not been their task or aim to give an analytic, syntactic, or semantic treatment of it. Not unsurprisingly, the phenomenon has recently been revived in more pragmatic and cognitively oriented approaches to

discourse semantics and in empirical studies. In this last section we briefly summarize some recent been results and findings.

Reinhard Blutner has developed a *DRT*-style account of presupposition in the framework of bi-directional optimality theory. Optimality theory has been a very successful paradigm in phonology and syntax (Prince & Smolensky 1997), and the bi-directional version of it has been very fruitful in semantics as well. In a bi-directional framework, optimalization strategies of both speakers and hearers are taken into account. When it comes to presupposition, put very bluntly, speakers ought not to say things that the hearers can figure it out themselves, but, also, speakers should say the things which they presuppose and which the hearers cannot be assumed to figure out for themselves.

Similar in spirit is a recent series of papers in which Philippe Schlenker has argued for a "transparant" pragmatic theory of presupposition, which is not built on a dynamic notion of meaning but on a bi-valent truth value assignment. Instead, Gricean-style pragmatics is invoked to deliver most of the kinds of presuppositions that Karttunen, Heim and Beaver have argued for (Schlenker 2008). A special issue of *Theoretical Linguistics* was devoted to proponents and opponents of this revolutionary or, as some claim, reactionary approach. Recently (Thomason et al.) proposed another computational perspective on presuppositions, in terms of "enlightened updates", dealing with commonsense reasoning about speakers' and hearers' intentions, acts, and plans.

Definitely theoretically oriented but much more empirically motivated, are recent studies conducted by Jennifer Spenader, Bart Geurts and Nausicaa Pouscoulous, and Emmanuel Chemla. Many discussions about presupposition, like many discussions about the semantics of expressions in natural language, get hampered by almost undecidable questions about natural language intuitions concerning complicated examples brought up from the philosopher's armchair, or the linguist's keyboard. Until recently, hardly any empirical tests had been made regarding these intuitions, but fortunately the aforementioned authors have started to fill this embarrassing gap. In the present state of affairs, however, it seems difficult to draw any specific conclusions, as the findings are not yet very decisive.

Other open questions, for which we are not going to claim an answer, and which are very much related, remain pretty open. There is still a debate about whether or not presupposition can be conceived of as a property of sentences, utterances, or speakers. Is there, after all, one notion of presupposition, or do the various so-called triggers get interpreted differently? Can we make do with a (partial) logic to interpret presuppositions, or do we have to implement a module in a (classical) system of interpretation that is especially designed to deal with presuppositions? An age-old question, first fully addressed by Gerald Gazdar, is: how do presupositions and implicatures relate? A final, old, consideration, the subject of current empirical research, is how presuppositions really project when they are embedded in modal and quantified structures. Obviously, it is also not so easy to gather empirical data, just as it is not so easy to gather analytical armchair data.

References

Aquinas, Thomas 2000–2009. Aristotle on Interpretation. In: *The Collected Works of St. Thomas Aquinas*, http://www.corpusthomisticum.org/: Corpus Thomisticum.

Aristotle 1975. *Categories and De Interpretation*. Oxford: Oxford University Press.

Bäuerle, Rainer, Uwe Reyle & Thomas Ede Zimmermann 2010. *Presuppositions and Discourse: Essays Offered to Hans Kamp*, volume 21 of CRISPI. Bingley, UK: Emerald Group.

Beaver, David 1995. Presupposition and Assertion in Dynamic Semantics. Ph.D. thesis, CCS, Edinburgh. Published in 2001 by CSLI Publications, Stanford.

Beaver, David 1997. Presupposition. In: Johan van Benthem & Alice ter Meulen (eds.) *Handbook of Logic and Language*, Amsterdam: Elsevier.

Bochvar, Dmitry Anatol'evich 1939. Ob odnom trehznachom iscislenii i ego primeneii k analizu paradoksov klassicskogo rassirennogo funkcional nogo iscislenija. *Matematiciskijs bornik* 4. English translation, 1981, "On a Three-valued Logical Calculus and its Applications to the Analysis of the Paradoxes of the Classical Extended Functional Calculus" in: *History and Philosophy of Logic* 2, pp. 87–112.

Dekker, Paul 2008. A Multi-Dimensional Treatment of Quantification in Extraordinary English. *Linguistics and Philosophy* 31(1), pp. 101–127.

Donnellan, Keith 1966. Reference and Definite Descriptions. *Philosophical Review* 75, pp. 281– 304.

Donnellan, Keith 1978. Speaker Reference, Descriptions and Anaphora. In: Peter Cole (ed.) *Syntax and Semantics*, Vol 9. Pragmatics, New York: Academic Press. pp. 47–68.

Fox, Danny 2008. Two Short Notes on Schlenker's Theory of Presupposition Projection *Theoretical Linguistics* 34(3), pp. 237–252.

Frege, Gottlob (ed.) 1884. *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner Verlag.

Frege, Gottlob 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* NF 100, pp. 25–50.

Frege, Gottlob 1969. *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*. Hamburg: Meiner Verlag. Edited by Hans Hermes and Friedrich Kambartel and Friedrich Kaulbach.

Gazdar, Gerald 1979. *Pragmatics. Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Geurts, Bart 1996. Local Satisfaction Guaranteed: A Presupposition Theory and its Problems. *Linguistics and Philosophy* 19(3), pp. 259–294.

Geurts, Bart 1999. *Presuppositions and Pronouns*, volume 3 of CRISPI. London: Elsevier.

Heim, Irene 1983. On the Projection Problem for Presuppositions. In: M. Barlow, D. Flickinger & M. Wescoat (eds.) *Proceedings of WCCFL II*. Stanford, California. Reprinted in Steven Davis (ed.) 1991, *Pragmatics: A Reader*, Oxford University Press.

Horn, Laurence R. 1985. Metalinguistic Negation and Pragmatic Ambiguity. *Language* 61, pp. 121–174.

Horn, Laurence R. 1996. Presupposition and Implicature. In: Shalom Lappin (ed.) *The Handbook of Contemporary Semantic Theory*, Oxford: Blackwell, pp. 299–319.

Kamp, Hans 2001. Presupposition Computation and Presupposition Justification: One Aspect of the Interpretation of Multi-Sentence Discourse. In: Myriam Bras & Laure Vieu (eds.) *Semantics and Pragmatic Issues in Discourse and Dialogue*, Amsterdam: Elsevier, pp. 57–84.

Karttunen, Lauri 1973. Presuppositions of Compound Sentences. *Linguistic Inquiry* 4, pp. 169–193. Karttunen, Lauri 1974. Presupposition and Linguistic Context. *Theoretical Linguistics* 1(1), pp. 181–193.

Karttunen, Lauri & Stanley Peters 1979. Conventional Implicature. In: Choon-Kyu Oh & David A. Dineen (eds.) *Syntax and Semantics 11—Presupposition*, New York: Academic Press. pp. 1–56.

Kempson, Ruth M. 1975. *Presupposition and the Delimitation of Semantics*. Cambridge: Cambridge University Press.

Kleene, Stephen Cole 1938. "On Notation for Ordinal Numbers".*The Journal of Symbolic Logic* 3(4), pp. 150–155.

Kleene, Stephen Cole 1952. *Introduction to Metamathematics*. Amsterdam: Notrh-Holland.

Kneale, William & Martha Kneale 1962. *The Development of Logic*. Oxford: Clarendon Press.

Laertius, Diogenes 1853. *The Lifes and Opinions of Eminent Philosophers*. http://classicpersuasion.org/pw/diogenes/

Łukasiewicz, Jan 1920. O logicetrójwartościowej. *Ruchfilozoficzny* 5, pp. 170–171. English translation, 1970, "On three-valued logic", in Ł. Borkowski (ed.), *Selected works by Jan Łukasiewicz*, NorthHolland, Amsterdam, 1970, pp. 8788.

Peters, Stanley 1977. A Truth-Functional Formulation of Karttunen's Account of Presupposition. In: Susan F. Schmerling & Carlota S. Smith (eds.) *Texas Linguistic Forum*, 6, University of Texas at Austin.

Prince, Alan & Paul Smolensky 1997. Optimality: From Neural Networks to Universal Grammar. *Science* 275, pp. 1604–1610.

van Rooij, Robert 2005. A Modal Analysis of Presupposition and Modal Subordination. *Journal of Semantics* 22(3), pp. 281–305.

van Rooij, Robert 2007. Strengthening conditional presuppositions. *Journal of Semantics* 24(3), pp. 289–304.

Russell, Bertrand 1957. Mr. Strawson on Referring. *Mind* 66, pp. 385–389.
van der Sandt, Rob A. 1988. *Context and Presupposition*. London: Croom Helm.
van der Sandt, Rob A. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics* 9(4), pp. 333–377.
Schlenker, Philippe 2008. Be Articulate: A Pragmatic Theory of Presupposition Projection. *Theoretical Linguistics* 34(3), pp. 157–212.
Seuren, Pieter A.M. 1985. *Discourse Semantics*. Oxford: Blackwell.
Seuren, Pieter A.M. 2005. Eubulides as a 20th-Century Semanticist. *Language Sciences* 27, pp. 75–95.
Soames, Scott 1989. Presupposition. In: Dov Gabbay & Franz Guenthner (eds.) *Handbook of Philosophical Logic*, Dordrecht: Kluwer, volume IV, Topics in the Philosophy of Language. pp. 553–616.
Stalnaker, Robert 1970. Pragmatics. *Synthese* 22, pp. 272–289. Reprinted in Donald Davidson and Gilbert Harman, 1972, *Semantics of Natural Language*, Reidel, Dordrecht, pp. 272–289.
Stalnaker, Robert 1978. Assertion. In: Peter Cole (ed.) *Syntax and Semantics 9—Pragmatics*, New York: Academic Press. pp. 315–332.
Strawson, Peter F. 1950. On referring. *Mind* 59, pp. 320–344.
Strawson, Peter F. 1952. *Introduction to Logical Theory*. London: Methuen & Co.
Thomason, Richmond H., Matthew Stone & David DeVault for Theoris. Enlightened Update: A computational architecture for presupposition and other pragmatic phenomena. In: Donna Byron, Craige Roberts & Scott Schwenter (eds.) *Presupposition Accommodation*, Ohio: Ohio State Pragmatic Initiative.
Zeevat, Henk 1992. Presupposition and Accommodation in Update Semantics. *Journal of Semantics* 9(4), pp. 379–412.

Further Reading

The reader may get a good sense of the wealth of discussion and examples relating to the topic of presupposition from the following overview papers and collections: (Soames 1989; Horn 1996; Beaver 1997; Bäuerle et al. 2010). A special issue of *Theoretical Linguistics*, 2008, Volume 34/3, has been devoted to the subject, as well as an ESSLLI workshop on "New Directions in the Theory of Presupposition" in Bordeaux, 2009, organized by Nathan Klinedinst and Daniel Rothschild (http://essllipresupposition2009.blogspot.com/

1.5

IMPLICATURE

Laurence Horn

A year after unveiling implicature in his William James lectures at Harvard, Paul Grice (1968) published one of the lectures, a study of the relation of word meaning to utterer's meaning situated within his overall framework for speaker meaning.

> The wider programme . . . arises out of a distinction I wish to make within the total signification of a remark, a distinction between what the speaker has *said* (in a certain favored and maybe in some degree artificial, sense of 'said'), and what he has 'implicated' (e.g., implied, indicated, suggested, etc.), taking into account the fact that what he has implicated may be either *conventionally* implicated (implicated by virtue of the meaning of some word or phrase which he has used) or *non-conventionally* implicated (in which case the specification of implicature falls outside the specification of the conventional meaning of the words used).
>
> (Grice 1968: 225; cf. also Grice 1989: 118)

Grice's footnote on the first page of this paper—"I hope that the material in this paper, revised and re-arranged, will form part of a book to be published by Harvard University Press"—is poignant to read, given that 21 years would elapse before this book would appear in print, and posthumously at that. Actually, the notions of conventional and non-conventional (specifically conversational) implicature, as defined here and as mapped by the familiar diagnostics of cancelability and detachability, had already been introduced several years earlier without the labels introduced above but already depicted as distinct varieties of (non-logical) implication (cf. Grice 1961: §3 and discussion in Horn 2004).

Conversational Implicature

Conversational implicature arises from the shared presumption that S and H are interacting to reach a shared goal. A speaker S saying *p* and implicating *q* counts on her interlocutor's ability to compute what was meant (by S's uttering *p* at a given point in the interaction) from what was said, based on the assumption that both S and H are rational agents. Speakers implicate, hearers infer; a hearer may recover an implicature, but to implicate is not to infer. Nevertheless, S's assumption that H will draw the appropriate inference is what makes implicature a rational expectation. The governing dictum is the Cooperative Principle: "Make your conversational contribution such as

is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange" (Grice [1967]1989: 26). This principle is instantiated by a set of general maxims of conversation whose exploitation potentially yields implicatures:

(1) QUALITY: Try to make your contribution one that is true.

1. Do not say what you believe to be false.
2. Do not say that for which you lack evidence.

QUANTITY:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

RELATION: Be relevant.
MANNER: Be perspicuous.

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief. (Avoid unnecessary prolixity.)
4. Be orderly.

Grice launched his characterization of implicature as an aspect of speaker meaning—what is meant without being said—at a moment when similar notions were in the air, especially the air of Oxford and its ordinary-language sphere of influence. Thus, a relation of "contextual implication" was invoked by Nowell-Smith (1954: 80–82) and evaluated by Hungerland (1960) in her eponymous paper: "A statement *p* contextually implies *q* if anyone who knew the normal conventions of the language would be entitled to infer *q* from *p* in the context in which they occur." But despite its name and context-relativity, such a relation differs substantively from implicature. The locus classicus for contextual implication is the context familiar from G.E. Moore: "When a speaker uses a sentence to make a statement, it is contextually implied that he believes it to be true" (Nowell-Smith 1954: 81). But the relation between (my stating) *He has gone out* and my believing that he has gone out cannot be assimilated to conversational implicature, for reasons given later by Grice himself:

> On my account, it will not be true that when I say that *p*, I conversationally implicate that I believe that *p;* for to suppose that I believe that *p* (or rather think of myself as believing that *p*) is just to suppose that I am observing the first maxim of Quality on this occasion. I think this consequence is intuitively acceptable; it is not a natural use of language to describe one who has said that *p* as having, for example, "implied", "indicated", or "suggested" that he believes that *p*. The natural thing to say is that he has expressed (or at least purported to express) the belief that *p*.
>
> (Grice 1989: 42)

Treating the Nowell-Smith/Hungerland relation as a proto-conversational implicature also instantiates three of Bach (2006)'s "Top ten misconceptions about implicature", #1 ("Sentences have implicatures"), #4 ("Gricean maxims apply only to implicatures"), and #6 ("All pragmatic implications are implicatures").

The difficulty of canceling such a (putative) implicature without epistemic or doxastic anomaly argues against such an analysis. The same applies to other cases of sincerity conditions Hungerland cites: e.g. the relation between *I promise to p* and *I intend to p*. Another of Nowell-Smith's examples of contextual implication does qualify as prefiguring Grice on the maxims, specifically Relation, although not implicature as such: "What a speaker says may be assumed to be relevant to the interests of the audience." This maxim may indeed be overridden, as Nowell-Smith and Hungerland both observe. But Hungerland (1960: 212) points out the heterogeneous nature of any construct that extends from this principle to the sincerity condition for assertions and promises; Nowell-Smith himself admits that the violation of the latter leads to "logical oddity," as in "It's raining, but I don't believe it is," while overriding the relevance rule chiefly runs the risk of boredom. As an anonymous reviewer notes, relevance violations may also yield confusion or signal a disparity of conversational goals; what they do not produce is the "logical oddity" of Moore's paradox.

In addition, conversational implicature differs from contextual implication (or nondemonstrative implication more generally) in being defined as a relation between a speaker (not a sentence) and a proposition and in arising typically from the exploitation of the maxims (Grice 1989: 26ff.; cf. Horn 2004); in the case of scalar implicature in particular, what is implicated depends on what isn't (but could have been) said. *Ceteris paribus*, a speaker uttering *Some F are* G implicates that (for all she knows) not all F are G because she would have been expected to express the stronger proposition if she had been in a position to do so. A speaker may opt for a weaker utterance from a belief that to express the latter might violate considerations of relevance, brevity, or politeness (cf. Horn 2004) but especially from a lack of certainty that the stronger counterpart holds.

Exactly a century before the William James lectures, John Stuart Mills offered these grounds for rejecting Sir William Hamilton's (1860) semantic account of *some* as "some only, some but not all":

> No shadow of justification is shown . . . for adopting into logic a mere *sous-entendu* of common conversation in its most unprecise form. If I say to any one, "I saw some of your children today", he might be justified in inferring that I did not see them all, not because the words mean it, but because, if I had seen them all, it is most likely that I should have said so: though even this cannot be presumed unless it is presupposed that I must have known whether the children I saw were all or not.
>
> (Mill 1867: 501)

The last point Mill makes in this passage, invoking the two-stage (epistemically governed) process involved in the computation of implicature recovery, is built into the rationality- and cooperation-driven Gricean model but is not directly captured in some current alternative theories; see Geurts (2010) for elaboration.

Mill's allusion to a tacit principle that requires the speaker to use the stronger *all* in place of the weaker *some* when possible, and invites the hearer to draw the corresponding inference when the stronger term is eschewed, is echoed by others in his own time—

> Whenever we think of the class as a whole, we should employ the term All; and therefore when we employ the term Some, it is implied that we are not

> thinking of the whole, but of a part as distinguished from the whole—that is, of a part only.
>
> (Monck 1881: 156)

– and in Grice's (e.g. Doyle 1951, Fogelin 1967: 20–22; see Horn 1990 and Chapman 2005: Chap. 5 for more on the chronology). Grice's contribution, beyond securing the naming rights to the relation in question, was to ground the operation of Mill's "sous-entendu[s] of common conversation" within an overall account of speaker meaning and the exploitation of tacit principles based on assumptions of the rationality and mutual goals of the interlocutors.

Mill's reasoning, codified by Grice in his first quantity maxim, is systematically exploited to yield upper-bounding SCALAR IMPLICATURES associated with relatively weak scalar operators, those configurable on a scale defined by unilateral entailment as in *<all, most, many, some>*. What is *said* in the use of a weak scalar value like those boldfaced in (2) is the lower bound *(. . . at least . . .)*; what is *implicated*, in the absence of contextual or linguistic cancellation, is the upper bound *(. . . at most . . .)*. What is *communicated*, *ceteris paribus*, is the 'two-sided reading' that combines what is said with what is implicated. Thus in (2d), "If we assert that a man who has acted in a particular way must be either a knave or a fool, we by no means assert . . . that he cannot be both" (Mill 1867: 512), but this unasserted 'exclusive' understanding of the disjunction may well be communicated.

(2)	ONE-SIDED READING →	TWO-SIDED READING
a. Pat has **3** children.	'at least 3'	'exactly 3'
b. You ate **some** of the cake.	'some if not all'	'some but not all'
c. It's **possible** she'll win.	'at least possible'	'possible but not certain'
d. He's a knave **or** a fool.	'. . . and perhaps both'	'. . . but not both'
e. It's **warm**.	'at least warm'	'warm but not hot'

The alternative view on which each scalar predication in (2) is lexically ambiguous between one-sided and two-sided readings falls afoul of Grice's (1989: 47) Modified Occam's Razor: "Senses are not to be multiplied beyond necessity." Scalar implicature was introduced and formalized in work by Horn (1972, 1989), Gazdar (1979), Hirschberg (1991), and Levinson (2000); cf. also Katzir (2007) and Geurts (2010) for insightful recent discussions of certain problems arising in the implementation of the central notions involved and Bontly (2005) for a defense of Modified Occam's Razor as a heuristic in acquisition.

One reflex of the centrality of scalar implicature for natural language is the systematic pattern of lexical gaps and asymmetries involving the negative incorporation in logical complexes. Thus, alongside the standard determiners *all, some, no,* we never find (in any language) a determiner **nall* ('not all'); alongside connectives corresponding to *and, or,* and sometimes *nor* (= 'and not'), we have no **nand* (= 'or not', 'not . . . and'). Although *some* does not contribute the same semantic content as *some not (not all)*, the use of either of the two values typically results in communicating the same information in a given context, 'some but not all'. The mutual quantity implicature holding between *some* and *some not* and the markedness of negation results in the superfluity of the latter for lexical realization (cf. Horn 1972, 1989 for details).

The implicature-based approach to scalar predications has been vigorously challenged by relevance theorists (see Carston 1988, 2002, 2004 and work reviewed therein), who take such sentences to involve not lexical but propositional ambiguity, with the pragmatically enriched two-sided meanings constituting not implicatures but EXPLICATURES, pragmatically derived components of propositional content. It has been suggested (Saul 2002) that some of the divergences between relevance theory (RT) and Gricean pragmatics are attributable to the focus of the former on the cognitive task of utterance interpretation as distinguished from the philosophical tradition in which Grice situated his own approach to implicature as a component of speaker meaning.

While the standard neo-Gricean view treats the cases in (2) homogeneously, many have questioned the tenability of this line for number words, as in (2a), arguing that such predications are semantically underspecified and contextually resolved, rather than assigned the weak, 'at least n' values by linguistic means. The underspecification line on cardinals, originally proposed by Carston (1988) and Koenig (1991), is extended in later work including Horn (1992), Geurts (1998), and Bultinck (2005). But while Ariel (2004) similarly disputes an implicature-based account of the upper bound of *most*-statements (the move from *most F are* G to 'not all F are G') based on a putative parallel between *most* and the cardinals regarding the status of the upper bound, there is evidence that *most* is distinct in behavior from the cardinals and that its meaning, like that of *some*, should be assigned the standard neo-Gricean account (unilateral semantics and upper-bounding scalar implicature) (cf. Horn 2006: §4). The strongest support for this hybrid view is the fact that a simple negative answer to a general scalar question, as opposed to one involving a cardinal value, always returns a 'less than' meaning, since this context selects descriptive and not metalinguistic negation. If you ask me whether most of the students passed, my negative response commits me to the proposition that 50 percent or fewer passed, not to the disjunction that either 50 percent-or-fewer passed or else all of them did. Yet it is just that disjunction that I must be asserting if my reply negates the proposition that '50–99 percent of the students passed' as it does on a semantic upper-bound account. On the other hand, if you asked me whether 10 students passed and I knew that 15 did, I must first determine whether you were asking me if at least 10 passed or exactly 10 passed before knowing whether to answer positively or negatively. A simple 'No' in response to 'Did 10 of the students pass?' might commit me to either 'Fewer than 10 passed' or 'Either fewer or more than 10 passed', depending on the context. Similarly, I would report that 'I am surprised that most of the students failed' only if I had expected at least half of them to pass, while my report that 'I am surprised that five students failed' is consistent with either a more pessimistic or more optimistic expectation.

In addition to such linguistic evidence, a now considerable body of empirical work indicates that the acquisition and processing of cardinals differs along a variety of parameters from that of other scalar values; note especially the studies in Papafragou and Musolino (2003), Hurewitz et al. (2006), and Breheny (2008). Thus, while acknowledging the important work on underspecification by Carston (and earlier related observations of Jay Atlas (1979)), we may need to distinguish the behavior of cardinals from that of their inexact scalar cousins. It is not obvious how the unitary explicature-based program for all scalar operators as in Carston (2002) can draw the necessary distinctions here, any more than the approach in Levinson (2000: 87–90) that retains the original (Horn 1972) implicature-based line for both cardinal and general scalar predications.

Two other challenges to the Gricean picture of implicatures involve the number and status of the maxims and the relationship between implicature and propositional content. To begin with the former issue, Grice himself later acknowledged (1989: 371ff.) that the four macroprinciples (inspired by Kant) and nine actual maxims in his inventory were somewhat overlapping and non-coordinate. The number of maxims has been revised both upward (Leech 1983) and downward. The dualistic program of Horn (1984, 1989, 2007a) begins by following Grice's lead (1989: 371) in ascribing a privileged status to Quality, on the grounds that without the observation of Quality, or Lewis's (1969) convention of truthfulness, no question of the observance of the other maxims can be considered (though see Sperber and Wilson 1986 for a dissenting view). The remaining maxims are subsumed under two countervailing functional principles whose role in communication and language change traces back to Paul (1889) and Zipf (1949). In this Manichaean model, implicatures may be generated by either the Q Principle (essentially 'Say enough', generalizing Grice's first submaxim of Quantity and collecting the first two 'clarity' submaxims of Manner) or the R Principle ('Don't say too much', subsuming Relation, the second Quantity submaxim, Relation, and Brevity).

The hearer-oriented Q Principle is a lower-bounding guarantee of the sufficiency of informative content, exploited to generate upper-bounding (typically scalar) implicata. The R Principle is a correlate of Zipf's principle of least effort dictating minimization of form, exploited to induce strengthening implicata; it is responsible for euphemism, indirect speech acts, neg-raising, and meaning change (Horn 2007a). In accord with the division of pragmatic labor (Horn 1984), a relatively unmarked form—briefer and/or more lexicalized—will tend to become R-associated with a particular unmarked, stereotypical meaning, use, or situation, while its periphrastic or less lexicalized counterpart, typically more complex or prolix, will tend to be Q-restricted by implicature to those situations outside the stereotype, for which the unmarked expression could not have been used appropriately (as in *kill* vs. *cause to die*, or *mother* vs. *father's wife*). Formalizations of the division of pragmatic labor have been undertaken within bidirectional optimality theory and game-theoretic pragmatics, although the results of such work are a matter of debate (cf. e.g. Blutner 2004, van Rooij 2009, and references cited therein).

Levinson's (2000) framework posits an interaction of Q, I (≈ Horn's R), and M (Manner) heuristics. Levinson's reconstruction of the division of pragmatic labor involves not Q but the M heuristic, given that *some* differs from *all* in informative content whereas *kill* differs from *cause to die* in complexity of production or processing. As Levinson concedes, however, the Q and M patterns are closely related, since each is negatively defined and linguistically motivated: S tacitly knows that H will infer from S's failure to use a more informative and/or briefer form that S was not in a position to have used that form. Unlike Q implicature, R/I-based implicature is not negative in character and is socially rather than linguistically motivated, typically yielding a culturally salient stereotype (Levinson 2000; cf. also Huang 2006 for a useful overview).

Relevance theorists (e.g. Sperber and Wilson 1986, Carston 2002) invoke one basic pragmatic principle, that of Relevance as defined in non-Gricean terms. It may be argued, however, that the RT program is itself covertly Manichaean, given that Relevance itself is calibrated as a minimax of effort and effect. As Carston (1995: 231) puts it, "Human cognitive activity is driven by the goal of maximizing relevance: that is . . . to derive as great a range of contextual effects as possible for the least expenditure of effort."

A clear departure from earlier Gricean models is in the move to reconceptualize implicature, and not just within RT. Over the last couple of decades, it has become a

truism within pragmatic theory that linguistically delivered aspects of meaning constitute a relative slender component of both the speaker-intended and the hearer-recovered message. The gap between what is literally expressed and what is communicated corresponds to aspects of meaning variously identified as conversational implicatures, conventional implicatures, presuppositions, implicitures, and explicatures. These are not simply alternate labels for a given phenomenon but different ways of conceptualizing the area between the said and the meant (or, for the hearer, between the said and the interpreted). Situating 'what is said' along this spectrum is itself controversial; what is said for Recanati (2001), Ariel (2008), and the relevance theorists is enriched by pragmatically derived material (hence constituting an explicature). Levinson, on the other hand, responds to the apparent need to accept "pragmatic intrusion" into what is said by allowing conversational implicatures to have truth-conditional consequences for the propositions in question, contra Grice; in cases like Deirdre Wilson's *It's better to meet the love of your life and get married than to get married and meet the love of your life*, an implicature ("P precedes Q") can feed into (rather than just being read off) what is said. (See Carston 2002 and Russell 2006 for illuminating discussions of the *and* case.)

For more orthodox Griceans, the pragmatically enriched proposition in such cases is distinct from what is said (see Horn 2009 for a defense of what Jenny Saul calls an "austere" conception of what is said). Bach (2001) observes that once we give up the "OSOP" (One Sentence, One Proposition) assumption, we can recognize that a given sentence may express not only more than one proposition but fewer. What is said in cases like *Dana is ready* may constitute not a truth-evaluable proposition but a propositional radical. Filling in such a radical or completing it within a given context to yield e.g. *Dana is ready to write a dissertation* yields not what is said (which is tightly constrained by the actual syntax) or an explicature (since there is nothing explicit about it), but rather an IMPLICITURE, a proposition implicit in what is said in a given context as opposed to a true implicature, a proposition read off what is said (or the way it is said). Similarly, scalar implicatures are reborn as scalar implic-i-tures. What Grice failed to recognize, argues Bach, is the non-exhaustive nature of the opposition between what is said and what is implicated.

Bach (2006) goes on to include "Scalar 'implicatures' are implicatures" as #9 in his hit parade of misconceptions: since a speaker uttering "Some of the boys went to the party" means not two separate things but just one thing, i.e. that some but not all of them went, this enriched proposition is an impliciture (built up from what is said), not an implicature (read off what is said). But it should be borne in mind that on the Gricean account (e.g. in Horn 1972, 1989), the scalar implicature here is "Not all of the boys went to the party"; this combines with what is said ("Some . . .") to yield what is communicated ("Some but not all . . ."). Thus the impliciture incorporates the scalar implicature rather than supplanting it.

Another of Bach's misconceptions, #2, is "Implicatures are inferences." While Levinson (2000) defines generalized conversational implicatures as DEFAULT INFERENCES, others argue that they are neither inferences—by definition an implicature is an aspect of speaker's meaning, not hearer's interpretation (cf. Bach 2001, Saul 2002)—nor true defaults. This last point is especially worth stressing in the light of much recent work in experimental pragmatics (see e.g. Noveck and Posada 2003, Bott and Noveck 2004, Breheny et al. 2006, Katsos 2008) suggesting that children and adults do not first automatically construct implicature-based enriched meanings for scalar predications and then, when the 'default' interpretation is seen to be inconsistent with the local context,

undo such meanings and revert to the minimal implicature-free meaning. To the extent that this empirical work on the processing of implicature recovery can be substantiated and extended, this is a very interesting result, but not (contrary to some claims) one that threatens the actual Gricean tradition, which predicts no automatic enrichment or default interpretation. This is clear from the passage distinguishing generalized and particularized implicature (Grice 1989: 37, italics added):

> I have so far considered only cases of what I might call 'particularized conversational implicature' . . . in which an implicature is carried by saying that p *on a particular occasion in virtue of special features of the context*, cases in which there is no room for the idea that an implicature of this sort is normally carried by saying that p. But there are cases of generalized conversational implicature. Sometimes one can say that the use of a certain form of words in an utterance would *normally (in the absence of special circumstances)* carry such-and-such an implicature or type of implicature.

The classic contrast at issue here dates back to Grice 1961: §3—the particularized implicature in the "Gricean letter of recommendation" for a philosophy job candidate *("Jones has beautiful handwriting and his English is grammatical")* vs. the generalized implicature in the logical disjunction (*"My wife is in Oxford or in London"*, implicating I don't know which). Crucially, an implicature may arise in an unmarked or default context without thereby constituting a default jor automatic inference. (See Bezuidenhout 2002, Jaszczolt 2005, and Geurts 2009 for more on defaults and their relation to implicature.)

Despite their substantial differences (from each other and from Grice) as to the role of implicature and the relation between what is implicated and what is said, the proponents of the approaches touched on above share Grice's commitment to situating implicature within a rationality-based pragmatics. On a competing view that has recently been elaborated by Chierchia (2004) and his colleagues, scalar implicatures in particular are generated locally as part of the grammar and/or the conventional lexical semantics of weak scalar operators. Support for this variety of defaultism involves an appeal to cases in which the Gricean model appears to yield the wrong results, thus arguing for local computation of "embedded implicatures." Others (e.g. Sauerland 2004, Russell 2006, Horn 2006) have challenged these conclusions and defended a global account of implicature along Gricean lines as both empirically and theoretically desirable. In particular, Geurts (2009, 2010) provides a broad survey of the landscape. Drawing a distinction between marked L[evinson]-type cases and unmarked C[herchia]-type cases of putative locality effects, Geurts (2009) argues that unlike the latter type, the Levinsonian contrast-induced narrowings represent true problems for a classical Gricean (or neo-Gricean) theory of implicature but shows that these can be handled by allowing upper-bounding to enter into the reinterpretation of what scalar operators express, a reinterpretation that is itself pragmatic in nature. In his treatise on Q-implicatures, Geurts (2010) argues that the conventionalist alternative to a Gricean approach is not only stipulative but also empirically flawed in failing to predict the full range of implicature-related results.

Conventional Implicature

In addition to introducing the very successful product line of conversational implicatures, Grice assembled an apparently inchoate class of phenomena under the rubric

of CONVENTIONAL IMPLICATURE, a non-cancelable but truth-conditionally transparent component of content whose success in the scholarly marketplace has been far more mixed. The coherence of this category has evoked much skepticism—Bach (1999) consigns it to the dustbin of mythology, Carston (2002: 134) remarks that "there simply is no such thing as 'conventional' implicature in relevance theory (or, we would argue, in reality)," while Potts (2005) undertakes a rehabilitation that redraws the descriptive map—but Grice's sketchy account of conventional content that does not affect the truth conditions of the asserted proposition has a rich lineage. Frege (1892, 1897, 1918) delineates a somewhat heterogeneous class of meanings that, while of linguistic interest, do not "affect the thought" or "touch what is true or false." While most recent scholarship follows Dummett (1973) in dismissing Frege's positive proposals in this area as representing a confused and subjective notion of "tone," this mischaracterizes Frege's actual account of the relevant phenomena. For a range of constructions including discourse particles *(but, even,* Ger. *ja, doch)*, subject-oriented adverbs, epithets, and other 'loaded' words, the approach favored by Frege and Grice remains eminently plausible (Barker 2003; Horn 2007b, 2008; Gutzmann 2008; Williamson 2009).

Further, such an approach can be fruitfully extended to a range of other natural language phenomena, including the familiar vs. formal second person singular ('T/V') pronouns of many modern European languages, evidential markers, and, arguably, the uniqueness/maximality condition on definite descriptions. In addition, certain syntactic constructions can be profitably analyzed along these lines such as the southern U.S. English "personal dative," a non-argument pronominal appearing in transitive clauses that obligatorily coindexes the subject as exemplified in *I love me some squid*, truth-conditionally equivalent to, but not fully synonymous with, 'I love squid' (Horn 2008). In each case, we find aspects of conventional content that are not entailed and do not fall inside the scope of logical operators.

The category of conventional implicature poses a complication for the distinction between the categories of what is said and what is meant. It could be argued that expressions falling under this analysis represent a recalcitrant residue for Grice (who was concerned with delineating what is said and what is conversationally, and hence calculably, implicated) as they did for Frege (who was concerned with the thought, i.e. with sense and potential reference); for both, detecting a conventional implicature facilitates the real work by clearing away the brush. But Grice did undertake to situate this relation within what we refer to (though he did not) as the "semantics/pragmatics divide." His contributions in this area, if not always accepted, are widely recognized, as in this passage from Davidson (1986: 161–62):

> It does not seem plausible that there is a strict rule fixing the occasions on which we should attach significance to the order in which conjoined sentences appear in a conjunction: the difference between 'They got married and had a child' and 'They had a child and got married.' Interpreters certainly can make these distinctions. But part of the burden of this paper is that much that they can do should not count as part of their linguistic competence. The contrast in which is meant of implied by the use of 'but' instead of 'and' seems to me another matter, since no amount of common sense unaccompanied by linguistic lore would enable an interpreter to figure it out. Paul Grice has done more than anyone else to bring these problems to our attention and help to sort them out.

But how does this sorting work? If descriptive content, reflecting what is said, is clearly semantic and if what is conversationally implicated (e.g. the 'for all I know, not both p and q' upper-bounding implicatum associated with the utterance of the disjunction *p or q* or the negative effect of the Gricean letter of recommendation) is pragmatic (pace Chierchia et al.), where is conventional implicature located? One standard view is that by falling outside what is said, it must be pragmatic (see e.g. Gutzmann 2008: 59). In Kaplan's words (1999: 20–21):

> According to Grice's quite plausible analysis of such logical particles as "but", "nevertheless", "although", and "in spite of the fact", they all have the same descriptive content as "and" and differ only in expressive content . . . The arguments I will present are meant to show that even accepting Grice's analysis, the logic is affected by the choice of particle . . . If this is correct, then generations of logic teachers, including myself, have been misleading the youth. Grice sides with the logic teachers, and though he regards the expressive content as *conventional* and hence (I would say) semantic (as opposed to being a consequence of his conversational maxims), he categorizes it with the maxim-generated *implicatures*.

To be sure, conventional implicatures are implicatures. But then again, they are conventional; we are indeed dealing here, unlike in the maxim-based cases, with aspects of content.

Two decades after the William James lectures, Grice revisited these categories in his Retrospective Epilogue (1989: 359–65), where he distinguished central and non-central modes of meaning through the criteria of FORMALITY ("whether or not the relevant signification is part of the conventional meaning of the signifying expression") and DICTIVENESS ("whether or not the relevant signification is part of what the signifying expression *says*"). Thus, when a speaker says "*p; on the other hand, q*" in the absence of any intended contrast of any kind between *p* and *q*, "one would be inclined to say that a condition conventionally signified by the presence of the phrase 'on the other hand' was in fact not realized and so that the speaker had done violence to the conventional meaning of, indeed had misused, the phrase 'on the other hand'." Crucially, however, "the nonrealization of this condition would also be regarded as insufficient to falsify the speaker's statement" (Grice 1989: 361). Thus, formality without dictiveness yields conventional implicature. (As for dictiveness without formality, a plausible candidate is the pragmatically enriched content of relevance theorists or the TRUTH-CONDITIONAL PRAGMATICS of Recanati 2001.)

In uttering a given sentence in a given context, the speaker may intend to communicate, and may succeed in communicating, more than one truth-evaluable proposition, but these communicated propositions do not necessarily have equal status; in *They are poor but happy*, the conjunctive content is truth-conditionally relevant, but the contrastive content is not. Yet *but* is not a synonym of *and*. Conventional implicatures constitute part of *encoded* content but not part of *truth-conditional* content per se; their falsity does not project as falsity of the expression to which they contribute (cf. Barker 2003). What they contribute is *use-conditional* meaning (Kaplan 1999, Gutzmann 2008).

Besides the defined properties of detachability and non-cancelability for such aspects of content (Grice 1961, 1989), additional diagnostics for conventional implicatures include their tendency to project out of embedded contexts and their immunity to

certain kinds of objection. In addition, conventional implicatures exhibit a contextual variability or DESCRIPTIVE INEFFABILITY (Potts 2007). Illustrations include the difficulty of pinning down the precise content of:

(3)
- the nature of the contrast/unexpectedness implicated by *but* (Dummett 1973)
- the characterization of the scalar conventional implicature associated with *even* (relative or absolute? unlikelihood or noteworthiness?)
- the nature of the expressive attitude embodied in racial and ethnic slurs and other epithets (Williamson 2003, 2009, Potts 2007)
- the precise notion of uniqueness or individuability associated with the use of definite descriptions (cf. Horn 2007b)
- the appropriateness implicatures for *tu* vs. *vous* or other **T** vs. **V** 2nd singular pronouns within a given context in a community of practice (**T** can be affectionate, presumptuous, comradely, or condescending; **V** can be polite, aloof, diplomatic, or hostile; cf. Brown and Gilman 1960)
- the implicatures of intentionality or positive affect for personal datives in Appalachian English *(He found/#lost him a laptop)* and related affectedness implicatures (benefactive or malefactive) for non-argument datives in a variety of languages (cf. Horn 2008)

Whence this property of ineffability? Modulo the recognized problems of vagueness, it is plausible that the edges of truth-conditional meaning should remain largely discrete, while inconsistency in the mental representation of non-truth-conditionally relevant content is less pernicious. If you know generally that my use of *vous* rather than *tu* signals some aspect of formal respect, distancing, and/or lack of intimacy, my precise motives can be left underdetermined, but if you don't know whether I'm using a second-person or third-person pronoun, the indeterminacy would be more costly. Similarly, it's crucial to know whether I bought the car for myself or for my son, and hence the identity of the referent for a true indirect object pronoun, but whether or not you can figure out precisely why I affirm "*I bought me a truck for my son*" rather than simply "*I bought a truck for my son*," no differences in argument structure or truth conditions emerge.

For Frege and Grice, identifying the class of conventional implicature-licensing constructions—including a motley collection of scalar particles, speaker-oriented sentence adverbs, epithets and slurs, prosodic features, evidential markers, "affected" pronominals, and word order effects—serves to characterize them in terms of what they are not: they do not affect the thought or the truth-conditionally relevant meaning of a given expression, and at the same time they are not derivable from general principles of rationality. While they share the former property with conversational implicatures, they differ crucially from them in the latter respect.

Since classical rhetoricians first recognized figures in which we say less and mean more ("*minus dicimus et plus significamus*"—Servius and Donatus, cited in Hoffmann 1987: 21), semanticists and pragmaticists have explored the domain of the said and the meant. In delineating and differentiating conversational and conventional implicatures, Grice provided a map for all further explorations of these boundary disputes, explorations that continue to inform our understanding of the nature of meaning in natural language.

Acknowledgments

I am grateful to Barbara Abbott, Kent Bach, Betty Birner, Robyn Carston, Gregory Ward, and the anonymous reviewer for Routledge for their helpful comments, pointers, and objections; the usual disclaimers apply. Thanks also to the editors for their patience.

Related Topics

1.2 Semantics and Pragmatics
1.6 Pragmatic Enrichment
1.7 Meaning and Communication
1.14 Relevance Theory.

References

Aikhenvald, A. (2004) *Evidentiality*. Oxford: Oxford University Press.
Ariel, M. (2004) *"Most," Language* 80: 658–706.
Ariel, M. (2008) *Pragmatics and Grammar*. Cambridge: Cambridge University Press.
Atlas, J. D. (1979) "How Linguistics Matters to Philosophy: Presupposition, Truth, and Meaning," in C.-K. Oh and D. Dinneen (eds.), *Syntax and Semantics 11: Presupposition*, 265–81. New York: Academic Press.
Bach, K. (1994) "Conversational Impliciture," *Mind and Language* 9: 124–62.
Bach, K. (1999) "The Myth of Conventional Implicature," *Linguistics and Philosophy* 22: 327–66.
Bach, K. (2001) "You Don't Say?" *Synthese* 127: 11–31.
Bach, K. (2006) "The Top 10 Misconceptions about Implicature. In B. Birner and G. Ward (eds.) *Drawing the Boundaries of Meaning*, 21–30. Amsterdam: John Benjamins.
Barker, S. (2003) "Truth and Conventional Implicature," *Mind* 112: 1–33.
Beaney, M. (ed., 1997) *The Frege Reader*. Oxford: Blackwell.
Bezuidenhout, A. (2002) "Generalized Conversational Implicatures and Default Pragmatic Inferences," in J. Campbell et al. (eds) *Meaning and Truth: Investigations in Philosophical Semantics*, 257–83. New York: Seven Bridges Press.
Blutner, R. (2004) "Pragmatics and the Lexicon," in Horn and Ward (eds.), 488–515.
Bontly, T. (2005) "Modified Occam's Razor: Parsimony Arguments and Pragmatic Explanations," *Mind & Language* 20: 288–312.
Bott, L. and I. Noveck (2004) "Some Utterances Are Underinformative: The Onset and Time Course of Scalar Inferences," *Journal of Memory and Language* 51: 437–57.
Breheny, R. (2008) "A New Look at the Semantics and Pragmatics of Numerically Quantified Noun Phrases," *Journal of Semantics* 25: 93–140.
Breheny, R., N. Katsos, and J. Williams (2006) "Are Generalized Scalar Implicatures Generated By Default? An On-Line Investigation Into The Role of Context in Generating Pragmatic Inferences," *Cognition* 100: 434–63.
Brown, R. and A. Gilman (1960) "The Pronouns of Power and Solidarity," in T. Sebeok (ed.) *Style in Language*, 253–77. Cambridge, MA: MIT Press.
Bultinck, B. (2005) *Numerous Meanings: The Meaning of English Cardinals and the Legacy of Paul Grice*. Oxford: Elsevier.
Carston, R. (1988) "Implicature, Explicature, and Truth-Conditional Semantics," in R. Kempson (ed.) *Mental Representations: The Interface Between Language and Reality*, 155–81. Cambridge: Cambridge University Press.
Carston, R. (1995) "Quantity Maxims and Generalized Implicatures," *Lingua* 96: 213–44.
Carston, R. (2002) *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
Carston, R. (2004) "Relevance Theory and the Saying-Implicating Distinction," in Horn and Ward (eds.), 633–56.
Chapman, S. (2005) *Paul Grice: Philosopher and Linguist*. Basingstoke: Palgrave Macmillan.
Chierchia, G. (2004) "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface," in A. Belletti (ed.) *Structures and Beyond*, 39–103. Oxford: Oxford University Press.
Chierchia, G., D. Fox, and B. Spector (forthcoming) "The Grammatical View of Scalar Implicatures and the

Relationship Between Semantics and Pragmatics", in C. Maienborn et al. (eds.) *Handbook of Semantics: An International Handbook*. Berlin: Mouton de Gruyter.
Davidson, D. (1986) "A Nice Derangement of Epitaphs," in R. Grandy and R. Warner (eds.) *Philosophical Grounds of Rationality*, 157–74.
Doyle, J. (1951) "In Defense of the Square of Opposition," *The New Scholasticism* 25: 367–96.
Dummett, M. (1973) *Frege: Philosophy of Language*. Oxford: Duckworth.
Fogelin, R. (1967) *Evidence and Meaning*. New York: Humanities Press.
Frege, G. (1892) "On *Sinn* and *Bedeutung*," Translated by M. Black, in Beaney (ed.) 151–71.
Frege, G. (1897) "Logic," unpublished paper, appearing in H. Hermes et al. (eds.) *Nachgelassene Schriften*, 139–61. Hamburg: Felix Meinen, 1969. Extract translated by P. Long and R. White, reprinted in Beaney (ed.) 227–50.
Frege, G. (1918–19) "Thought," translated by P. Geach and R. H. Stoothoff, in Beaney (ed.) 325–45.
Gazdar, G. (1979) *Pragmatics*. New York: Academic Press.
Geurts, B. (1998) "Scalars," in P. Ludewig and B. Geurts (eds.) *Lexikalische Semantik aus kognitiver Sicht*, 95–117. Tübingen: Gunter Narr Verlag.
Geurts, B. (2009) "Scalar Implicature and Local Pragmatics," *Mind and Language* 24: 51–79.
Geurts, B. (2010) *Quantity Implicatures*. Cambridge: Cambridge University Press.
Grice, H. P. (1961) "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, Supplementary Volume 35: 121–52.
Grice, H. P. (1967) "Logic and Conversation," William James lectures delivered at Harvard University.
Grice, H. P. (1968) "Utterer's Meaning, Sentence-Meaning, and Word-Meaning," *Foundations of Language* 4: 225–42.
Grice, H. P. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
Gutzmann, D. (2008) *On the Interaction Between Modal Particles and Sentence Mood in German*. Mainz: German Institute, Johannes Gutenberg University.
Hamilton, Sir W., of Edinburgh (1860) *Lectures on Logic, Volume I*. Edinburgh: Blackwood.
Hirschberg, J. (1991) *A Theory of Scalar Implicature*. New York: Garland.
Hoffmann, M. (1987) *Negatio Contrarii: A Study of Latin Litotes*. Assen: Van Gorcum.
Horn, L. (1972) *On the Semantic Properties of Logical Operators in English*. PhD dissertation, UCLA. Distributed by Indiana University Linguistics Club, 1976.
Horn, L. (1984) "Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature," in D. Schiffrin (ed.) *Meaning, Form, and Use in Context (GURT '84)*, 11–42. Washington, D.C.: Georgetown University Press.
Horn, L. (1989) *A Natural History of Negation*. Chicago: University of Chicago Press. (Reissued with a new introduction, Stanford: CSLI, 2001.)
Horn, L. (1990) "Hamburgers and Truth: Why Gricean Inference is Gricean," *BLS 16*, 454–71.
Horn, L. (1992) "The Said and the Unsaid," *SALT II*, 163–92. Columbus: Ohio State University, Department of Linguistics.
Horn, L. (2004) "Implicature," in Horn and Ward (eds.) 3–28.
Horn, L. (2006) "The Border Wars: a Neo-Gricean Perspective," in K. Turner and K. von Heusinger (eds.) *Where Semantics Meets Pragmatics*, 21–48. Oxford: Elsevier.
Horn, L. (2007a) "Neo-Gricean Pragmatics: a Manichaean Manifesto," in N. Burton-Roberts (ed.) *Pragmatics*, 158–83. Basingstoke: Palgrave Macmillan.
Horn, L. (2007b) "Toward a Fregean Pragmatics: *Voraussetzung, Nebengedanke, Andeutung*," in I. Kecskes and L. Horn (eds.) *Explorations in Pragmatics: Linguistic, Cognitive, and Intercultural Aspects*, 39–69. Berlin: Mouton de Gruyter.
Horn, L. (2008) "'I love me some him': The Landscape of Non-Argument Datives," in O. Bonami and P. Cabredo Hofherr (eds.) *Empirical Issues in Syntax and Semantics 7*: 169–92. Downloadable at http://www.cssp.cnrs.fr/eiss7
Horn, L. (2009) "WJ-40: Implicature, Truth, and Meaning," *International Review of Pragmatics* 1: 3–34.
Horn, L. and G. Ward (eds., 2004) *The Handbook of Pragmatics*. Oxford: Blackwell.
Huang, Y. (2006) *Pragmatics*. Oxford: Oxford University Press.
Hungerland, I. (1960) "Contextual Implication," *Inquiry* 3: 211–58.
Hurewitz, F., A. Papafragou, L. Gleitman and R. Gelman (2006) "Asymmetries in the Acquisition of Numbers and Quantifiers," *Language Learning and Development* 2: 77–96.
Jaszczolt, K. (2005) *Default Semantics: Foundations of a Compositional Theory of Acts of Communication*. Oxford: Oxford University Press.

Kaplan, D. (1999) "The Meaning of *Ouch* and *Oops:* Explorations in the Theory of *Meaning as Use*", unpublished ms., UCLA.
Katsos, N. (2008) "The Semantics/Pragmatics Interface from an Experimental Perspective: The Case of Scalar Implicature," *Synthese* 165: 385–401.
Katzir, R. (2007) "Structurally Defined Alternatives," *Linguistics and Philosophy* 30: 669–690.
Koenig, J.-P. (1991) "Scalar Predicates and Negation: Punctual Semantics and Interval Interpretations," *CLS 27*, 140–55.
Leech, G. (1983) *Principles of Pragmatics*. London: Longman.
Levinson, S. C. (2000) *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.
Lewis, D. (1969) *Convention*. Cambridge, MA: Harvard University Press.
Mill, J. S. (1867) *An Examination of Sir William Hamilton's Philosophy* (3d edn). London: Longman.
Monck, W. H. S. (1881) *Sir William Hamilton*. London: Sampson, Low.
Noveck, I. and A. Posada (2003) "Characterizing the Time Course of an Implicature: An Evoked Potentials Study," *Brain and Language* 85: 203–10.
Nowell-Smith, P. H. (1954) *Ethics*. Harmondsworth: Pelican Books.
Papafragou, A. and J. Musolino (2003) "Scalar Implicatures: Experiments at the Semantics-Pragmatics Interface," *Cognition* 86: 253–82.
Paul, H. (1889) *Principles of the History of Language*, translated by H.A. Strong. London: Macmillan.
Potts, C. (2005) *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
Potts, C. (2007) "The Expressive Dimension," *Theoretical Linguistics* 33: 165–98.
Recanati, F. (2001) "What is Said," *Synthese* 128: 75–91.
van Rooij, R. (2009) "Optimality-theoretic and Game-theoretic Approaches to Implicature," in E. Zalta (ed.) *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2009/entries/implicature-optimality-games/
Russell, B. (2006) "Against Grammatical Computation of Scalar Implicatures," *Journal of Semantics* 23: 361–82.
Sauerland, U. (2004) "Scalar Implicatures in Complex Sentences," *Linguistics and Philosophy* 27: 367–91.
Saul, J. (2002) "What is Said and Psychological Reality: Grice's Project and Relevance Theorists' Criticisms," *Linguistics and Philosophy* 25: 347–72.
Sperber, D. and D. Wilson (1986) *Relevance: Communication and Cognition*. Oxford: Blackwell.
Williamson, T. (2003) "Understanding and Inference," *Aristotelian Society Supp.* 77: 249–93.
Williamson, T. (2009) "Reference, Inference, and the Semantics of Pejoratives," in J. Almog and P. Leonardi (eds.) *Philosophy of David Kaplan*, 137–59. Oxford: Oxford University Press.
Wilson, D. and D. Sperber (2004) "Relevance Theory," in Horn and Ward (eds.) 607–32.
Zipf, G. K. (1949) *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

1.6

PRAGMATIC ENRICHMENT

François Recanati

1 Pragmatic Enrichment and Conversational Implicature

The phrase "pragmatic enrichment" is sometimes used in a broad sense to refer to the process by virtue of which the content conveyed by an utterance comes to include all sorts of elements which are contextually implied without being part of what the utterance literally means. In his class notes "Pragmatic Enrichment: Introduction by Examples," Chris Potts (2007) gives examples like the following:

> (1) John and Mary have recently started going together. Valentino is Mary's ex-boyfriend. One evening, John asks Mary, "Have you seen Valentino this week?" Mary answers, "Valentino's been sick with mononucleosis for the past two weeks." Valentino has in fact been sick with mononucleosis for the past two weeks, but it is also the case that Mary had a date with Valentino the night before.

Mary's utterance clearly suggests a negative (and false) answer to the question: "Have you seen Valentino this week?" Literally, however, she only says that Valentino has been sick with mononucleosis for the past two weeks, and that is true.

In his William James Lectures delivered at Harvard in the late 1960s, Grice described the mechanism through which one can mean one thing by saying another (Grice 1989). The term coined by Grice for that mechanism is "conversational implicature." Thus, in (1), Mary 'says' that Valentino has been sick with mononucleosis for the past two weeks, and 'conversationally implicates' that she has not seen him this week.

On Grice's account, implicatures are derived through an inference which enables the interpreter to grasp the speaker's communicative intention despite the fact that it's not articulated in words. The inference makes use of two crucial premises (in addition to background knowledge): (1) the fact that the speaker has said what she has said (here, that Valentino has been sick etc.), and (2) the fact that the speaker, qua communicator, obeys (or is supposed to obey) the rules of communication or, as Grice puts it, the "maxims of conversation." Because of the role played by the first premise in the inference through which they are derived, conversational implicatures belong to the 'post-semantic' layer of interpretation. This means that, in order to derive the implicatures of an utterance, an interpreter must start by identifying what the speaker literally says.

(The properly *semantic* layer of interpretation corresponds to the identification of what is said—the proposition literally expressed.)

It is customary in the literature to equate 'pragmatic' and 'post-semantic,' but this is not quite right, for two reasons. First, there is a 'pre-semantic' layer of interpretation which, like the post-semantic layer, involves a good deal of pragmatic reasoning aimed at figuring out the speaker's communicative intention. Semantic interpretation takes *disambiguated* expressions as inputs, and computes their contents as a function of the contents of their constituents and the way they are syntactically combined (plus the context, when the sentence contains indexicals or context-sensitive expressions). No such computation can take place unless the sentence has been properly disambiguated. In this pre-semantic task, background knowledge and the presumption that the speaker obeys the maxims of conversation play an obvious role. As David Kaplan points out, if a haberdasher says "I am out of checks" to a banker, "whether the utterance takes place in the store or at the bank will help the banker to determine what the haberdasher has said" (Kaplan 1978: 229), and it will do so by pre-semantically selecting the relevant reading for the word "check." Pragmatics is therefore involved at the pre-semantic level, just as it is involved at the post-semantic level.

Pragmatics is also involved at the *semantic* level, and that is the second reason why one should reject the equation of 'pragmatic' and 'post-semantic.' To compute the proposition expressed by an utterance, it is necessary to assign contextual values to indexicals, (unbound) pronouns etc. For example, if the speaker uses a demonstrative pronoun and says, "'She is cute," the hearer must determine who the speaker means by "she" in order to fix the utterance's truth-conditional content. This contextual assignment of values to indexicals and pronouns is a pragmatic affair, just like disambiguation or the derivation of implicatures. In all three cases one has to use contextual cues (and the presumption that the speaker obeys the maxims of conversation) to figure out what the speaker's intentions are.

To sum up, utterance interpretation involves pragmatic reasoning at three levels: (i) at the pre-semantic level one must determine which (disambiguated) sentence is intended among various candidates compatible with the proferred signal; (ii) at the semantic level one must assign contextual values to the context-sensitive elements in the sentence in order to determine the proposition expressed; (iii) at the post-semantic level one must derive the conversational implicatures. Only in the last type of case (the implicatures) does the inference presuppose that the proposition expressed (what is said) has been identified.

Now, *pace* Potts, the phrase 'pragmatic enrichment' on its standard use is meant to *contrast* with 'conversational implicature,' instead of being a non-technical synonym for it (as it is for Potts). On the standard use, while conversational implicatures belong to the post-semantic layer of interpretation, pragmatic enrichment belongs to the semantic layer: It affects the proposition expressed by an utterance. This puts pragmatic enrichment in the same ballpark as the assignment of contextual values to indexicals and free pronouns.

2 Pragmatic Enrichment and Semantic Under-specification

What, then, is the difference between pragmatic enrichment and the contextual assignment of values to indexicals and pronouns if both belong to the semantic layer of interpretation? Following Kaplan, let us call 'content' the contribution an expression

makes to the truth-conditions of the utterance in which it occurs. Context-sensitive expressions are such that their linguistic meaning does not fix their content: their content depends upon the context. In the case of standard indexicals, pronouns etc., the expression is like a free variable to which a content must be *assigned* in context, and its linguistic meaning merely constrains the contents which may be contextually assigned to it. For example, the linguistic meaning of the pronoun "she" in the above example carries gender and number features which (presuppositionally) constrain the reference of the pronoun: the reference must be a female individual. The content of the pronoun is whatever is assigned to it in the context of utterance (in virtue of the speaker's intentions etc.), *provided* the assigned value satisfies the presuppositions carried by the pronoun.

In other cases, a contextual value is not directly assigned to the expression, but the content of the expression still depends upon the context. For example, 'ready' and 'tall,' without an accompanying prepositional phrase, are context-sensitive because they can be used to ascribe different properties in different contexts—the property of being *ready for the walk* vs *ready for dinner*, or the property of being *tall for a sixth-grader* vs. *tall for a basketball player*. Some authors talk of pragmatic enrichment in that sort of case: the meaning of 'ready' or 'tall,' they say, is under-specified and has to be pragmatically enriched through the contextual provision of a meaning constituent expressible by means of a prepositional phrase ('for dinner,' 'for a sixth-grader'). Since the enrichment in question is mandatory, the meaning of the semantically under-specified item can be represented as involving a free variable to which a value must be contextually assigned: thus 'ready' arguably means something like *ready for x* and 'tall' means something like *tall for an X*. In the syntax, however, there arguably is no variable—only a semantically under-specified lexical item whose meaning carries what Kratzer (2009) has called a "conceptual" variable. The variable is introduced in the semantics, via the lexical entry for the item, while in the case of pronouns and indexicals, the lexical item occurring in the syntactic structure is itself (like) a variable.

The treatment of context-sensitive expressions like 'ready' and 'tall' I have just described is actually controversial. Some authors think that, associated with lexical items like 'ready' and 'tall,' there *is* a (covert) variable/pronominal expression in the syntax (Stanley 2000, Marti 2006). In context, a value must be assigned to that covert variable, just as a contextual value must be assigned to overt pronouns. If this is true, then there is no need to talk of semantic under-specification or pragmatic enrichment in that sort of case—it is just an ordinary case of contextual assignment of value to pronominal elements in syntactic structure.

In other cases, however, the idea of semantic under-specification triggering pragmatic enrichment may be harder to resist. Thus Kratzer (2010) says that "pragmatic enrichment is not just allowed, but required, for the interpretation of noun-noun compounds like *swan boat*." Neither 'swan' nor 'boat' is context-sensitive (let us assume)—both lexical items have a content independent of context—but the compound 'swan boat' is: its content must be pragmatically fixed through the contextual provision of a relevant relation *R* between swans and boats, such that only boats bearing *that relation* to swans end up in the extension of 'swan boat.' What we seem to have here is constructional context-sensitivity: the context-dependence of the content of 'swan boat' is not inherited from the context-dependence of the contents of its parts, but comes from the construction itself. The content of the noun–noun construction is semantically under-specified: it has to be fleshed out through the contextual provision of the relevant relation *R*.

To be sure, there have been attempts to provide 'indexical' analyses of compound nouns (CNs). Thus, according to Weiskopf (2007: 175),

> The linguistic meaning of CNs contains a phonologically unrealized open variable or otherwise indexical expression that picks out the relation among their constituents, and this relation is filled in by pragmatic mechanisms that draw on features of the context of utterance.

But the variable R* which Weiskopf posits is introduced via the semantic rules associated with the noun–noun construction, which he calls the "rules of CN Interpretation," e.g.:

$$[[N1\ N2]] = \lambda x[\mathbf{N'}2(x)\&R^*(x,\mathbf{N'}1)]$$

So it is, as he puts it, "the meaning of CNs" which carries the variable. The variable in question is therefore a conceptual variable and is not supposed to be present in the syntax. (In this respect it is somewhat misleading to say that the variable is "phonologically unrealized," for that notion only applies to elements present in syntactic structure.)

3 Free Pragmatic Enrichment

I have just illustrated the use of the phrase 'pragmatic enrichment' in connection with semantic under-specification and conceptual variables. On that use it is possible to say, as Kratzer does, that pragmatic enrichment is required for the interpretation of a given type of expression. But that is not the standard use of the phrase 'pragmatic enrichment.' On the standard use, pragmatic enrichment is *never* required. It is the standard use of the phrase (and the phenomenon it designates) that I will focus on in what follows.

On the standard use, just as it contrasts with conversational implicatures, pragmatic enrichment contrasts with the assignment of values to free variables, *whether the variables in question are construed as conceptual (introduced via lexical entries) or syntactic*. What characterizes the contextual assignment of values to free variables of all categories is that it is *mandatory*: no proposition is expressed unless a value is assigned to the variable. Thus, if I say "John is ready," but the context does not answer the question "Ready for what?", I haven't said anything definite; I haven't expressed a complete proposition. Similarly, if I say "she is cute," but no individual can be contextually singled out as the reference of the demonstrative pronoun 'she', no proposition has been expressed. Likewise, if I use the phrase 'swan boat', but the context does not specify any relevant relation *R* between swans and boats, my utterance fails to express a complete content. In all these cases, where the provision of some contextual element is mandatory, I talk of 'saturation': a slot has to be contextually filled, leaving the utterance semantically incomplete if it remains unfilled. In contrast, pragmatic enrichment—or, more explicitly, *free* pragmatic enrichment—is not mandatory but optional. There is enrichment in that sense only when the following three conditions are jointly satisfied: (i) the context adds some element to the interpretation of the utterance; (ii) that element is truth-conditionally relevant—it affects the proposition expressed (so this is unlike conversational implicatures); yet (iii) its contextual provision is not necessary, in the sense that if that element was left aside, the utterance would still express a complete proposition (this is in contrast to indexicals, pronouns, and the other type of context-sensitive expressions mentioned above).

Consider the following example from Robyn Carston:

(2) Jean went to Austria last week and she ran into John.

One naturally interprets the second conjunct ("she ran into John") as expressing the proposition that Jean ran into John in Austria. The location of the event of running into John is contextually assumed to be the same as the location mentioned in the first conjunct, but that assumption is defeasible, as contextual assumptions typically are. What is important, however, is that the contextual provision of a location for the event mentioned in the second conjunct is not mandatory: it is perfectly correct to interpret a sentence like "she ran into John" as saying that Jean ran into John in some place or other, without the place in question being contextually specified or specifiable. This is in contrast with standard indexicals or pronouns: For the sentence does not express a complete proposition unless the indexicals or unbound pronouns it contains are assigned definite values.

4 Pragmatic Modulation in General

Free enrichment is only one type of case in which a pragmatic process that is optional (in contrast to the contextual assignment of values to indexicals and free variables of all sorts) nevertheless affects the proposition expressed. In *Literal Meaning* and subsequent work, I used the term 'modulation' (as opposed to 'saturation') as a generic term for the entire class. The general idea is this:

> In context the meaning of words is adjusted or 'modulated' so as to fit what is being talked about. Sense modulation is essential to speech, because we use a (more or less) fixed stock of lexemes to talk about an indefinite variety of things, situations and experiences. Through the interaction between the context-independent meanings of our words and the particulars of the situation talked about, contextualised, modulated senses emerge, appropriate to the situation at hand.
> (Recanati 2004: 131)

Saturation is a pragmatic process of contextual value-assignment that is triggered (and made obligatory) by something in the sentence itself, namely, the linguistic expression to which a value is contextually assigned. The expression itself acts as a variable in need of contextual instantiation. So pragmatics comes into play (in order to determine which value to assign in context to the variable), but it does so under the guidance of the linguistic material: the pragmatic process of saturation is a 'bottom-up' process in the sense that it is signal-driven, not context-driven. In contrast, modulation is a 'top-down' or 'free' pragmatic process, i.e. a process which is not triggered by an expression in the sentence but is context-driven and takes place for purely pragmatic reasons—in order to make sense of what the speaker is saying.

Modulation operates on the meaning of an expression (simple or complex) and returns a meaning of the same type. In free enrichment, the output sense is more specific than the input sense. Free enrichment typically works by narrowing down the extension of an expression through the contextual addition of a component to its meaning. For example, through free enrichment, the (intransitive) verb 'to smoke' can come to mean *to smoke marijuana*, if the context makes it clear that that is what we are talking

about. What is added to the meaning of the verb is a specification of its internal accusative—the type of thing smoked—just as in (2) what is added is a specification of the location of the event of running into John.

Sense extension (also called 'loosening' or 'broadening') is another modulation process which works in the opposite direction: The meaning of an expression is made *less* specific, and its extension typically wider, through the contextual deletion of a component of its meaning (Cohen 2002: chapters 3 and 5; Franks 1995). Thus consider:

(3) The ATM swallowed my credit card

This can be given a literal reading, if we imagine a context à la Putnam in which ATMs turn out to be living organisms (for only living organisms can swallow). But the sentence can also and typically will be interpreted non-literally. In an ordinary context, 'swallow' will be given an extended sense, corresponding to what ATMs sometimes do with credit cards (something which, indeed, resembles swallowing). The sentence may be true, on such a reading, even though no real swallowing takes place.

Another key modulation process is predicate transfer (Nunberg 1995). In predicate transfer the input sense is a property F, and the output sense is another property F', definable in terms of F, along the following lines:

$$F' =_{\text{def}} \lambda x\, [(Qy\text{: } Rx,y)\ Fy]$$

In this schema representing the transferred property F', "*Q*" and "*R*" are placeholders for, respectively, a quantificational determiner and a relation that is left free (within certain limits) and will be determined in context. So F' is the property one has whenever, for *Q* object *y* one bears relation *R* to, *y* is *F*. In Nunberg's famous example, "I am parked out back," F is the property of being parked out back (a property only a vehicle can have), *R* is the ownership relation between cars and car owners, and *Q* may be the existential quantifier. F', then, is the property of having a car that is parked out back. It is that property which the words 'parked out back' arguably ascribe to the speaker in "I am parked out back."

5 The 'Syntactic' Approach

Even though free enrichment is only one type of modulation among others, the phrase 'pragmatic enrichment' is often used as a generic term, synonymous with my own term 'pragmatic modulation'. One reason for this is that modulation can be thought of as operating on *expressions* ('syntactic' construal), rather than directly on their meanings ('semantic' construal). When so construed, modulation amounts to *supplementing* the overt expression with implicit elements, and so it counts as a form of 'enrichment', whatever its semantic effect.

On the semantic construal of pragmatic modulation, the meaning of an expression is mapped to a distinct meaning. On the syntactic construal, it is the expression (or a corresponding mental representation) that is mapped to a more complex expression/representation through the addition of extra elements. For example, a sentence like (4) has several readings:

(4) There is a lion in the middle of the piazza

On one reading 'lion' is given a non-literal interpretation through predicate transfer and (4) may be true even if, literally, there is no lion in the middle of the piazza (but only a statue of a lion). An advocate of the 'syntactic' approach to pragmatic modulation will argue that 'lion' is mapped to 'representation of a lion' through the addition of an operator REPRESENTATION OF. Likewise, sense extension will be said to proceed through the addition of some kind of similarity operator. On this view pragmatic modulation always is a form of enrichment or supplementation of a (linguistic or mental) representation, making it into a more complex representation.

The syntactic approach to pragmatic modulation has one putative advantage over the semantic approach:It can provide a simple response to an argument against pragmatic enrichment known as the 'binding argument'. According to the argument, implicit elements of content (e.g. the implicit reference to Austria in the second conjunct of (2)) cannot, in general, be due to free pragmatic enrichment, because the elements in question can be made to vary with the cases introduced by a quantifier, as in

(5) Whenever I light a cigarette, it starts raining.

On the most salient reading, this means that whenever I light a cigarette, it starts raining *at the location where I light the cigarette*. Here the location of the raining event is an implicit element of content, as is the location of the event of running into John in example (2) above, but it varies with the cases introduced by "whenever I light a cigarette." Implicit bound readings like this raise a problem for pragmatic enrichment accounts, Stanley (2000) argued, because binding requires a variable that can be bound (here, a location variable), and variables are syntactic objects. According to free enrichment accounts of implicit content, however, the location in (2), (5) and similar examples is supposed to be inarticulated in the syntax, which means that there is no location variable in the syntax which could be bound; nor could the variable be generated through pragmatic enrichment, since pragmatic enrichment (on the semantic construal) only maps meanings to meanings. However, assuming that this is correct and that implicit bound readings cannot be generated through pragmatic enrichment semantically construed, they can still be generated through pragmatic enrichment syntactically construed. Nothing prevents the syntactic-enrichment process from contextually supplementing the overt sentence with implicit variables or pronouns directly or indirectly bound by the overt quantifier (Recanati 2002: 338–42).

6 Two Variants of the 'Syntactic' Approach

There are two main variants of the 'syntactic' approach to pragmatic modulation: the natural-language (NL) variant and the language-of-thought (LOT) variant.

On the NL variant, pragmatic enrichment is the process whereby a natural language expression (that which is actually uttered, e.g. the word 'lion' in the above example) is mapped to another natural language expression ('lion statue'), which is not uttered but might have been, and which can be contextually recovered from the sentence that is actually uttered. This is the 'shorthand' view put forward (inter alia) by Kent Bach (1994, 2000) and Stephen Neale (2000, 2004). On this view modulation works by enriching or supplementing the sentence with contextually recoverable extra materials, as in the following examples, from Kent Bach:

Jack and Jill went up the hill *[together]*.
Jack and Jill are married *[to each other]*.
Jill got married and *[then]* got pregnant.
You'll succeed if *[and only if]* you work hard.
There are *[approximately]* 30 students per class.
France is hexagonal *[roughly speaking]*.
Otto has *[exactly]* three cars.
Felix has always been an honest judge *[since he's been a judge]*.
Adele hasn't had lunch *[today]*.
You're not going to die *[from this cut]*.

As Bach (2000: 262–63) puts it, "we generally speak loosely, omitting words that could have made what we meant more explicit and letting our audience fill in the gaps." In such cases "what the speaker means is a qualified version of what he is saying. He would have to utter the words in brackets (or roughly equivalent words—the exact words don't matter) to make what he meant more explicit." The words in question "are *not* part of the sentence (these are not cases of syntactic ellipsis) but the speaker could have uttered them if he deemed that necessary."

On the LOT variant, put forward by relevance theorists (Sperber and Wilson 1986, Carston 2002) and by Jerry Fodor (2001), natural language expressions are assumed to translate into language of thought strings (sometimes called 'semantic' or 'conceptual' representations). Pragmatic enrichment is the process whereby the language of thought string into which the uttered expression translates is contextually enriched with extra symbols, to which nothing corresponds in the uttered expression. It is the richer language of thought string resulting from the enrichment process which determines the (intuitive) truth-conditions of the utterance: the truth-conditions of the utterance are the truth-conditions one gets by subjecting the contextually enriched LOT string (rather than the natural language sentence or its mentalese translation) to model-theoretic interpretation.

As Paul Elbourne points out, the NL variant is more constrained than the LOT variant. "Enrichment in [the first kind of] theory can only yield meanings that could be obtained by syntactically building on the material present in the syntax [of the natural language expression]. No such constraint is present in the [other kind of] theory" (Elbourne 2008: 95). "Theories of [the second] kind, then, are potentially vulnerable in the following way: if a reading turns out to be unavailable for a given sentence, they cannot avail themselves of syntactic constraints to explain its absence, . . . and they will face real difficulty if the reading in question can be argued to be pragmatically plausible" (Elbourne 2010). Still, it is easy to introduce an analogous constraint into the LOT variant, and it is that sort of hybrid view which Stephen Neale seems to be defending. On such a hybrid view a LOT string is considered as an admissible enrichment of (the mentalese translation of) the uttered sentence only if it is the mentalese translation of a sentence one can get by supplementing the uttered sentence with extra materials.

Even if we take the syntactic constraint on board, there is an obvious overgeneration objection looming. Many readings one could get by freely enriching the uttered sentence with extra words are simply not available (Stanley 2002, 2005, Elbourne 2010). Extra constraints will therefore be needed to filter out the absent readings. (See Elbourne 2008 and Hall 2008 for attempts to address some of the overgeneration worries.) Overgeneration is also a problem for semantic approaches to pragmatic

modulation. In general, pragmatic modulation/enrichment is such a powerful mechanism that it needs to be suitably constrained. Thus predicate transfer as a general mechanism ought to make it possible to interpret the sentence "I was once driven by Jean Gabin" to mean that the speaker owns a car that was once driven by Jean Gabin. That reading is very unlikely, however, for reasons that are essentially pragmatic and are well-worth investigating (see Nunberg 1979, 1995 for a pragmatic approach to the constraints on predicate transfer).

7 Is Pragmatic Enrichment Relevant to Semantics?

It is tempting to appeal to pragmatic enrichment/modulation whenever an element of meaning, e.g. the location of the event of running into John in (2), remains implicit yet seems to be truth-conditionally relevant. The standard strategy in linguistics consists in dealing with such examples by positing unpronounced elements in the syntax. But in many cases, such covert elements can only be detected by pragmatic means, and there is only a short step from the idea that they are detected on a pragmatic basis to the idea that they are generated on a pragmatic basis. As Ariel Cohen (2009) puts it, if an element is phonologically null, the hearer can't hear it; if the hearer nevertheless proceeds as if that element were present, s/he must apply some reinterpretation of the input so as to introduce that element. Modulation is the case in which such reinterpretation proceeds on a purely pragmatic basis.

Despite the appeal of the idea, and the fact that it promises to simplify both syntax and semantics, there is deep resistance to it and it has been rather underutilized in formal semantics. In the final sections of this entry, I discuss what I take to be the main sources of resistance, besides the overgeneration worry which has already been mentioned.

First and foremost, there is resistance to the very idea that free pragmatic processes might affect 'what is said'—the proposition expressed by an utterance—as suggested by the idea that modulation belongs to the semantic layer of interpretation. Free pragmatic processes are pragmatic processes that are not mandated by the linguistic material but take place for wholly pragmatic reasons, in order to make sense of the speaker. They are 'top down' processes, rather than 'bottom up' pragmatic processes triggered by something in the linguistic signal. That such processes could affect what is said seems to be a contradiction in terms. 'What is said', the truth-conditional content of an utterance, is what is *literally* said, and that—by definition—has to be determined by the conventions of the language. Pragmatics can enter the picture, provided its role is to assign a contextual value to a lexical item in a bottom-up manner, i.e. in accord with (and under the guidance of) the conventional meaning of that context-sensitive item. In contrast, what King and Stanley (2005: 118–19) call "strong pragmatic effects," i.e. pragmatic effects achieved in order to make sense of the speech act without being linguistically mandated, take us into the realm of speaker's meaning, away from literal meaning.

Insofar as this argument is based upon a certain understanding of the phrase 'what is said' (or 'what is literally said'), it is *not* substantive, but verbal. There is no doubt that one can *define* 'what is said' in such a way that only weak pragmatic effects can affect what is said. But what the advocate of pragmatic modulation means by 'what is said' corresponds to the *intuitive truth-conditional content* of the utterance. Now, the intuitive truth-conditions of an utterance of (3) or (4) above *are* affected by free pragmatic processes of predicate transfer or sense extension: (4) can be intuitively true even though no real lion stands in the middle of the piazza, and (3) can be true even though no

real swallowing takes place. So there is a sense of 'what is said' in which what is said is affected by pragmatic modulation. Assuming this is true, this does not prevent us from defining another notion of what is said, conforming to literalist standards.

The second objection to pragmatic modulation (or rather, to the idea that it affects semantic content) is the following. If free pragmatic processes are allowed to affect semantic content, semantic content leaps out of control—it is no longer determined by the rules of the language but varies freely, à la Humpty Dumpty. But then, how can we account for the success of communication? Communication (content sharing) becomes a miracle since there is nothing to ensure that communicators and their addressees will converge on the same content. Now communication *is* possible (it takes place all the time), and there is no miracle. It follows that we should give up the view that free pragmatic processes play a role in the determination of semantic content (Cappelen and Lepore 2005: chapter 8).

This argument fails, I believe, because the problem it raises is a problem for everybody. *Whenever* the semantic value of a linguistic expression (e.g. the reference of a demonstrative) must be pragmatically inferred, the question arises: What guarantees that the hearer will be able to latch on to the exact same semantic value as the speaker? Whether the pragmatic process at stake is saturation or modulation is irrelevant as far as this issue is concerned, so the argument fails as an argument specifically intended to cast doubt on pragmatic modulation.

8 The Systematicity Objection

The most important source of resistance to the idea of pragmatic modulation is the fear that if something like pragmatic modulation is allowed as a determinant of semantic content, the project of constructing a systematic, truth-conditional semantics for natural language will be doomed to failure. Here is a first sketch of an argument for that conclusion:

> In contrast to the contextual assignment of values to indexicals, modulation is not driven by the linguistic meaning of words. Nothing in the linguistic meaning of the words whose sense is modulated tells us that modulation ought to take place. Modulation takes place purely as a matter of context, of 'pragmatics'; what drives it is the urge to make sense of what the speaker is saying. So modulation *is* unsystematic. If we allow it as a determinant of semantic content, we make it impossible to construct a systematic theory of semantic content.

Thus formulated, the argument is not convincing. Let us grant the objector that modulation is unsystematic. Still, it is easy to make room for it within a systematic semantics (Pagin and Pelletier 2007, Westerståhl forthcoming). In general, nothing prevents unsystematic factors from being handled systematically, by being assigned their proper place in the theory. In the case at hand, assuming the semantic approach to pragmatic modulation, we can define a function *mod* taking as arguments an expression *e* and the context *c* in which it occurs: the value of *mod* is the particular modulation function *g* that is contextually salient/relevant/appropriate for the interpretation of that expression in that context. If no modulation is contextually appropriate and the expression receives its literal interpretation, the value of *mod* will be the identity function. In this framework, we can distinguish between the *literal* sense of a simple expression *e*, namely

its semantic interpretation *I(e)*, and the *modulated* sense $M(e)_c$ carried by an occurrence of *e* in context *c*. The modulated sense of an expression *e* (in context *c*) results from applying the contextually appropriate modulation function *mod (e, c) = g* to its semantic interpretation *I(e)*:

$$M(e)_c = mod\ (e, c)\ (I(e)) = g\ (I(e))$$

So far, this is very standard: in distinguishing *I(e)* from $M(e)_c$ we are just appealing to the traditional semantics/pragmatics distinction. What is *not* standard is the claim that **the semantic interpretation of a *complex* expression (e.g. a sentence) is a function of the *modulated senses* of its parts and the way they are put together** (Recanati 2010: chapter 1). This is what examples like (3) and (4) suggest if we take at face value the effects of modulation on truth-conditional content which they seem to display. On the resulting view the semantic process of composition and the pragmatic process of sense modulation are intertwined. For simple expressions their semantic interpretation is their literal sense, but *for complex expressions pragmatic modulation is allowed to enter into the determination of semantic content*. This is non-standard, certainly, but there is nothing unsystematic about this view.

The systematicity objection can be understood differently, however. In introducing modulation (in contrast to saturation), I said that in saturation "pragmatics comes into play, but does so under the guidance of the linguistic material," whereas modulation "is not triggered by an expression in the sentence but takes place for purely pragmatic reasons—in order to make sense of what the speaker is saying." This suggests that, insofar as it involves modulation, utterance interpretation does not significantly differ from "the kind [of interpretation] involved in interpreting kicks under the table and taps on the shoulder" (Stanley 2000: 396). Stanley objects that "we do not interpret these latter sorts of acts by applying highly specific rules to structured representations," as formal semanticists say we do in the linguistic case. Accepting the semantic relevance of pragmatic modulation thus seems incompatible with the program of formal semantics, according to which utterance interpretation is a systematic affair and differs, in this respect, from the (pragmatic) interpretation of gestures.

Thus understood, however, I think the objection is confused. Even though free pragmatic processes, i.e. pragmatic processes that are not mandated by the standing meaning of any expression in the sentence, are allowed to enter into the determination of truth-conditional content, still, in the framework I have sketched, *they come into the picture as part of the compositional machinery*. Semantic interpretation remains *grammar-driven* even if, in the course of semantic interpretation, pragmatics is appealed to, not only to assign contextual values to indexicals and free variables but also to freely modulate the senses of the constituents in a top-down manner. Semantic interpretation is *still* a matter of determining the sense of the whole as a function of the (possibly modulated) senses of the parts and the way they are put together.

Acknowledgment

The research leading to this paper has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013) / ERC grant agreement no 229 441—CCC.

References

Bach, K. (1994) "Conversational Impliciture," *Mind and Language* 9: 124–62.
Bach, K. (2000) "Quantification, Qualification and Context," *Mind and Language* 15: 262–83.
Cappelen, H. and Lepore, E. (2005) *Insensitive Semantics*, Oxford: Blackwell.
Carston, R. (2002) *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.
Cohen, A. (2009) "No Quantification without Reinterpretation," Workshop Genericity: Interpretation and Uses, Paris, May 11–13, 2009.
Cohen, L. J. (2002) *Knowledge and Language: Selected Essays*, Dordrecht: Kluwer.
Elbourne, P. (2008) "The Argument from Binding," *Philosophical Perspectives* 22: 89–110.
Elbourne, P. (2010) "Incomplete Descriptions and Sloppy Identity," Contextualism and Compositionality workshop, Paris, May 17–19 2010.
Fodor, J. (2001) "Language, Thought and Compositionality," *Mind and Language* 16: 1–15.
Franks, B. (1995) "Sense Generation: A 'Quasi-Classical' Approach to Concepts and Concept Combination," *Cognitive Science* 19: 441–505.
Grice, P. (1989) *Studies in the Way of Words*, Cambridge, MA.: Harvard University Press.
Hall, A. (2008) "Free Enrichment or Hidden Indexicals?" *Mind and Language* 23: 426–56.
Kaplan, D. (1978) "DTHAT," *Syntax and Semantics* 9: 221–43.
King, J. and Stanley, J. (2005) "Semantics, Pragmatics, and the Role of Semantic Content," in Z. Szabo (ed.) *Semantics versus Pragmatics*, Oxford: Clarendon Press, 111–64.
Kratzer, A. (2009) *Context and Modality*, The Context and Content Lectures, Paris, October–December 2009.
Kratzer, A. (2010) Abstract for "Syntactic Constraints on Meaning Composition: Deriving the Meanings of Conditionals," Contextualism and Compositionality workshop, Paris, May 17–19, 2010.
Marti, L. (2006) "Unarticulated Constituents Revisited," *Linguistics and Philosophy* 29: 135–66.
Neale, S. (2000) "On being Explicit: Comments on Stanley and Szabo, and on Bach," *Mind and Language* 15: 284–94.
Neale, S. (2004) "This, That and the Other," in M. Reimer and A. Bezuidenhout (eds.), *Descriptions and Beyond*, Oxford: Clarendon Press, 68–182.
Nunberg, G. (1979) "The Non-Uniqueness of Semantic Solutions: Polysemy," *Linguistics and Philosophy* 3: 143–84.
Nunberg, G. (1995) "Transfers of Meaning," *Journal of Semantics* 12: 109–32.
Pagin, P. and Pelletier, J. (2007) "Context, Content and Communication," in G. Preyer and G. Peter (eds.) *Context-Sensitivity and Semantic Minimalism*, Oxford: Clarendon Press, 25–62.
Potts, C. (2007) "Pragmatic Enrichment: Introduction by Examples," class notes, University of Massachusetts, Amherst, Fall 2007.
Recanati, F. (2002) "Unarticulated Constituents," *Linguistics and Philosophy* 25: 299–345.
Recanati, F. (2004) *Literal Meaning*, Cambridge: Cambridge University Press.
Recanati, F. (2010) *Truth-Conditional Pragmatics*, Oxford: Oxford University Press.
Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*, Oxford: Blackwell.
Stanley, J. (2000) "Context and Logical Form," *Linguistics and Philosophy* 23: 391–434.
Stanley, J. (2002) "Making It Articulated," *Mind and Language* 17: 149–68.
Stanley, J. (2005) "Semantics in Context," in G. Preyer and G.Peter (eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, Oxford: Clarendon Press, 221–53.
Weiskopf, D. (2007) "Compound Nominals, Context, and Compositionality," *Synthese* 156: 161–204.
Westerståhl, D. (forthcoming) "Compositionality in Kaplan Style Semantics," in W. Hinzen, E. Machery, and M. Werning (eds.) *The Handbook of Compositionality*, Oxford: Oxford University Press.

1.7

MEANING AND COMMUNICATION

Kent Bach

"If I didn't know any words, you wouldn't know what I mean." This astute observation, made by my granddaughter Sophia when she was four, might suggest that knowing what a speaker's words mean is all it takes to know what she means in using them and, that communicating is just a matter of putting one's thoughts into words. Sophia didn't suggest that and, indeed, the theme of this chapter is that communication is more complicated than that. For even if you know what my words mean, you might not know what I mean.

Introduction

Words mean things, speakers mean things in using words, and these need not be the same. For example, if you say to someone who has just finished eating a super giant burrito at the Taqueria Guadalajara, "You are what you eat," you probably do not mean that the person is a super giant burrito. So we need to distinguish the meaning of a linguistic expression—a word, phrase, or sentence—from what a person means in using it. To simplify matters, let us pretend that an utterance is always of a sentence (and, for mnemonic purposes, let our imagined speaker be a *s*he and *h*earer be a *h*e).

This chapter is concerned with the relationship between linguistic meaning and what speakers mean in using language. It will not take a stand on the nature of linguistic meaning itself, a difficult question on which there are many views, some of which are discussed elsewhere in this volume. However, we will assume that it is one thing for a sentence to have a certain meaning (or meanings, if it ambiguous) and another for a speaker to mean something, whether the same thing or something else, in using it. This leaves open whether what words mean in a language ultimately comes down to what speakers mean by them, as argued by Grice (1968) and by Schiffer (1972). We will assume also that speakers ordinarily, as members of the same linguistic community, share knowledge of the meanings of the expressions they use. This is part of their linguistic knowledge, which also includes knowledge of phonological (or orthographic) form and syntactic structure. What matters for us is that linguistic knowledge is only part of the knowledge that people bring to bear when they communicate with one another. Our examples will only begin to illustrate how, mostly for efficiency's sake but sometimes for other reasons, people commonly try to communicate more than what their sentences mean, and often even manage to make themselves understood.

We will explore one particularly ingenious idea about this, due to Grice (1957). He thought that communication involves a special sort of intention on the part of the speaker and that successful communication involves a special sort of inference on the part of the intended audience. In using a sentence to try to communicate something, a speaker has an audience-directed intention that is in a certain way self-referential. Specifically, the speaker intends the listener to figure out what the speaker means partly on the supposition *that* the speaker intends him to do so. The hearer's job is to figure out what the speaker means, partly on the basis *that* he is intended to do so. This is possible because unlike intentions in general, a communicative intention is one whose fulfillment consists simply in its recognition. In zeroing in on what this involves, we will need to keep in mind that people generally do not use sentences merely to communicate but primarily to affect one another in various ways.

Linguistic Meaning and Speaker Meaning

What we mean is generally connected, though sometimes only remotely, to what our words mean. To appreciate this, consider a case where there is no such connection at all. Suppose to gain entrance into a private club you must utter a three-word sentence whose words begin, respectively, with "a," "b," and "c." You say, "Always be cool," and you are let in. Clearly the meanings of your words are irrelevant to what you mean ("I'm a member—let me in"). You could just as well have said, "Antibodies battle chlamydia." But this is an exceptional case. Ordinarily the meanings of the words you use do matter. Nevertheless, their meanings do not determine what you mean in using them. There are various ways in which this can be.

First of all, this can be because of *ambiguity*. A sentence can have more than one meaning because it contains an ambiguous expression, like 'bar' in (1),

(1) Because of his excessive drinking, Judge Jones was banned from the bar.

or because it is structurally ambiguous, like (2),

(2) The chicken is ready to eat.

Because of the lexical ambiguity in (1), a speaker, though being literal, could mean either that Judge Jones was banned from legal practice or that he was banned from a certain drinking establishment. Similarly, a literal speaker of the structurally ambiguous (2) could be talking about either a hungry chicken or a broiled chicken. In each case what the speaker means corresponds to only one of the things the sentence means.

Another way linguistic meaning can fail to determine what the speaker means is via *nonliterality*. For instance, although sentence (4) means something analogous to what (3) means,

(3) Farmer Frank was up to his ears in mud.
(4) Farmer Frank was up to his ears in debt.

a speaker is likely to mean something quite different. What he means is related to but distinct from the linguistic meaning, since he means that Farmer Frank was only figuratively up to his ears in debt. In other cases, involving *indirection*, a speaker means what

the sentence means (or one of the things it means, if it is ambiguous) but means something else as well. If a friend asks you for something to drink and you utter (5),

(5) There's some beer in the fridge.

presumably you mean not only that there is some beer there but also that your friend may help himself to some of it. Finally, an utterance can be both nonliteral and indirect, as in a likely case of a mother saying (6) to her slightly sadistic son,

(6) I'm sure Felix likes having his tail pulled.

She means not only that the cat doesn't like having his tail pulled but also that her son should stop pulling it. These and similar examples (see Bach and Harnish 1979: chapter 4) illustrate different ways in which what the speaker means can depart from what the sentence means. The speaker may mean one of the things the sentence means, as with (1) or (2), something quite distinct from anything it means, as with (4), or both, as with (5).

There is a minor complication here. We need to distinguish what a speaker means *by* an expression and what she means *in using* it. This distinction is evident from (7), for example,

(7) Dr. Frankenstein's lavatory blew up.

where the speaker probably means laboratory by "lavatory." This distinction is also needed to handle utterances of ambiguous sentences like (1) and (2), where what a speaker means by her sentence does not include each of the things it means. In these cases what the speaker means by the sentence determines which meaning is operative in her utterance of them. With (4), however, there is no linguistic ambiguity. Despite what a speaker would mean *in using* the words "up to his ears," she does not really mean anything different *by* them than she would in uttering (3). The phrase seems not to be ambiguous but rather to have two uses, one literal and one figurative, one corresponding to its single meaning and the other a derivative one. The speaker is exploiting the single (literal) meaning of her words in order to mean something else in using them. With (5) the speaker means both what the sentence means and something else as well. When she utters "There's some beer in the fridge," she means *by* those words just what they mean (not quite, actually, since they do not specify the fridge in question). Yet *in* uttering (5) she means more than just that, namely that the hearer may help himself to some beer.

Communicative Intentions

Intuitively, to mean something in uttering a sentence is to intend to communicate something to one's audience. But what is it for an intention to be communicative? In his groundbreaking article "Meaning," Grice (1957) offered an original answer to this question. He observed that meaning something (trying to communicate it) is not simply a matter of thinking something and acting with the intention of somehow causing one's audience to entertain that thought. After all, one's intention could be covert. You might, for example, make self-deprecating remarks intending to get people to think you are modest. They might think that but certainly not if they recognize your intention.

Nor is it enough that one's intention be overt. Say you point to a cracked window with the intention of getting someone to believe that the window is broken. Seeing that it is, they will come to believe that but not by virtue of recognizing your intention.

Grice's idea was that communicative intentions are intentionally overt and that this feature plays a special role in their fulfillment. That is, in trying to communicate something to others by saying something, a speaker intends the audience to recognize that intention partly by taking into account that they are so intended. As Grice characterized the distinctively self-referential or "reflexive" character of communicative intentions, a speaker means something by his utterance only if he intends his utterance "to produce some effect in an audience by means of the recognition of this intention" (Grice 1957/1989: 220). Not just any sort of effect will do, and later we will consider just what sort of "effect" this is. But first we need to appreciate Grice's basic idea.

To get a feel for it, consider what goes on in the following games, which have something in common with linguistic communication. Take the game of charades, in which one player uses gestures and other bodily movements to help the second player identify what she has in mind. The first player has a self-referential intention, for part of what she intends is for the second player to take into account the very fact that she intends her gestures etc. to enable him to figure out what she has in mind. Nothing like this goes on in the game of 20 questions, where the second player uses answers to yes-or-no questions to narrow down the possibilities of what the first player has in mind. Here the only cooperation required is honest answers on the part of the first player. Like charades, simple games of tacit coordination, first discussed by Schelling (1960: 54–58), also involve self-referential intentions. The first player selects and records an item in a certain specified category, such as a letter of the alphabet, a liquid, a mode of transportation, a city, or a U.S. president; the second player has one chance to guess it. In this game either both win or both lose. Both win if and only if the second player guesses right without any help. But what counts as guessing right? That depends entirely on what the first player has in mind, and that in turn depends entirely on what she thinks the second player, taking into account that she wants him to guess right, will think she wants him to guess. The second player guesses whatever he thinks she wants him to guess. To appreciate how this cooperative guessing game works, play this game with a friend. Try additional categories too, and consider why some work better than others.

When players use the above categories, they usually both pick the letter 'A', water, cars, the city in which they are located, and the current president. Each 'correct' choice stands out in some salient way from other members of the category. Grice's idea was in effect that successful communication involves something of the same sort. In uttering a sentence, a speaker has something in mind that she is trying to convey and intends the hearer to figure out what that is; in hearing the speaker utter the sentence, the hearer tries to figure out what the speaker intends to convey, partly on the basis of being so intended. That is, the hearer is to take into account that he is intended to figure out what the speaker intends to convey. It is the meaning of the words uttered, of course, that provides the primary input (along with the fact that the speaker uttered them, presumably with the intention to communicate something), but what they mean does not determine what the speaker means. Even if what she means is precisely what her words mean, the fact that she is speaking literally is not determined by what they mean—she could have meant something else. What is loosely called "context" plays a key role here, not in determining what the speaker means—that is a matter of the speaker's communicative intention—but of enabling the hearer to ascertain (a different sense of "determine")

what the speaker means. Context comprises whatever other considerations the hearer is to take into account in so doing. It is information that is mutually available to the speaker and the hearer, information that the speaker intends the hearer to take into account in figuring out what the speaker means (see Bach 2005).

Communication and Speech Acts

If Grice was right, there is something distinctively self-referential about a communicative intention. But what is it that the speaker intends? What sort of effect does she intend to produce on her audience, partly "by means of the recognition of this intention"? We cannot take up this question until we reckon with the fact that utterances, though generally communicative, are not made just with communicative intentions—they are not *merely* acts of communication. As first investigated by Austin (1962) in his aptly titled *How to Do Things with Words*, a *speech act* is a multi-layered affair. In this respect they are no different from most intentional actions. Moving one's arm in a certain way can, given the right intentions and circumstances, also be a case of pushing away a second plate of pasta, of sticking to one's diet, and of trying to impress one's spouse. Notice that this is not a series of actions but, rather, a single bodily movement comprising a multiplicity of nested actions. The same occurs when one utters a sentence.

Utterances as Three-level Speech Acts

Austin distinguished three distinct levels beyond the mere act of uttering a sentence. There is the act *of* saying something, what one does *in* saying it, and what one does *by* saying it. He dubs these, respectively, *locutionary*, *illocutionary*, and *perlocutionary* acts. Saying isn't just a matter of meaning what one's words mean. The locutionary act is, as Austin defined it (1962: 95), the act of using a sequence of words "as belonging to a certain vocabulary . . . and as conforming to a certain grammar, . . . with a certain more or less definite sense and reference" (this qualification is needed to allow for resolving any ambiguity and fixing any indexical reference). Importantly, Austin did not mean the act of uttering particular words, reported with direct quotation as for example in (8):

(8) Bernanke says, "Inflation is not a problem."

The locutionary act is reported, rather, with indirect quotation:

(9) Bernanke says that inflation is not a problem.

Bernanke did not have to use the words, "Inflation is not a problem," to say that inflation is not a problem. He did not even have to be speaking in English. Regardless of the means by which he said that, or even if he had uttered something else, say "Inflation is not on the horizon," he could have performed the illocutionary act of assuring the public that prices and interest rates won't go up significantly, thereby performing the perlocutionary act of assuaging at least one of their economic fears.

The sentence uttered does not in general determine the type of illocutionary act being performed. Just as we can do one of several different things in shaking hands—introduce ourselves, greet each other, seal a deal, or bid farewell—so we can use a given sentence in various ways. For example, (10) can be used, depending on who is speaking

to whom and with what intention, as a prediction, a warning, a promise, a threat, or even an order.

(10) The riot squad will break up the demonstration.

It is partly, but only partly, because of what the sentence means that it has its various (literal) uses. After all, the sentence expresses, at least relative to a given context of utterance, a proposition about what a certain riot squad will do regarding a certain demonstration. However, the meaning of the sentence does not determine whether it is being used to predict, warn, promise, threaten, or order. And it certainly does not determine the perlocutionary effect which, depending on the illocutionary act being performed, could be anticipation, dispersal, assurance, fear, or action.

Communicative vs. Perlocutionary Intentions

Now we can spell out the difference, corresponding to the distinction between illocutionary and perlocutionary acts, between a speaker's communicative intention and her further intention in uttering a sentence. Intuitively, an act of communication, linguistic or otherwise, is an act of expressing oneself. This rather vague idea can be made more precise if we get more specific about what is being expressed. Take the case of an apology. If you utter, "[I'm] sorry I smashed your car," and intend this as an apology, you are expressing regret, in this case for smashing the hearer's car. Indeed, it seems that an apology just is the act of (verbally) expressing regret for something one did that adversely affected the hearer. It is communicative because it is intended to be taken as expressing a certain attitude, in this case regret (for smashing the hearer's car). That is only the communicative aspect of an apology. When you apologize, you may intend not just to express regret but also, hoping the hearer thinks you're sincere, to seek forgiveness. Seeking forgiveness is distinct from expressing regret, even though in saying that you're sorry for smashing the hearer's car you are doing both.

In general, speech acts are not merely acts of saying something and not merely communicative, illocutionary acts of expressing an attitude. They are also perlocutionary acts, performed to produce some effect on the audience. However, since the intended perlocutionary effect of a given type of illocutionary act can vary, it makes sense to distinguish different types of speech acts primarily by their illocutionary type, such as asserting, requesting, promising, and apologizing, which in turn may be distinguished by the type of attitude expressed. The perlocutionary act is generally, though not always, an attempt to get the hearer to form some correlative attitude, as this table illustrates.

Illocutionary Act	*Attitude Expressed*	*Intended Hearer Attitude*
statement	belief that p	belief that p
request	desire for H to D	intention to D
promise	firm intention to D	belief that S will D
apology	regret for D-ing	forgiveness of S for D-ing

These acts exemplify the four main categories of communicative illocutionary acts, which Bach and Harnish (1979: chapter 3), borrowing partly from the terminology of both Austin (1962) and Searle (1969), call *constatives, directives, commissives*, and *acknowledgments*.

If a type of act can be distinguished by type of expressed attitude, then an act of that type is communicationally successful if the hearer recognizes the attitude being expressed, such as a belief in the case of a statement and a desire in the case of a request, along with its content (what is believed, desired, or whatever). Any further perlocutionary effect the act has on the hearer, such inducing a belief or an action, or even just being taken as sincere, is not essential to its being a statement or a request and is not necessary for its success as an act of communication. It need not be sincere—the speaker might not actually possess the attitude she is expressing—and the hearer might not take her to be sincere. But there is no question about the speaker being sincere possessing the communicative intention itself, for this intention must be identified before the question of her sincerity can even arise. One can be unsuccessful in conveying one's communicative intention—by being too vague, ambiguous, or metaphorical, or even by being wrongly taken literally—but not insincere about it.

Self-Referential Intentions Again

Now we are in a position to fit Grice's idea of self-referential intentions, the key to his conception of communication, into the broader framework of speech act theory. This will enable us to pin down the sort of effect a speaker can intend "to produce in an audience by means of the recognition of this intention," and also to see why there is nothing paradoxical or mysterious about such an intention.

The Intended "Effect"

As Strawson (1964: 459) first pointed out, the relevant effect is understanding, rather than, as Searle (1969: 47) added, any further, perlocutionary effect on the hearer. And understanding, or what Austin (1962) called "uptake," is a matter of identifying what attitude (and its content) the speaker is expressing. We can think of meaning something as intending just such an effect: "to express an attitude is reflexively to intend the hearer to take one's utterance as reason to think one has that attitude" (Bach and Harnish 1979: 15). This formulation respects the difference between expressing an attitude and actually possessing it, not to mention having a further intention toward the hearer (Siebel 2003 expresses some worries about this formulation). Whatever else a speaker may be doing in performing a speech act, her communicative illocutionary act is just the act of expressing that attitude. Communicating successfully, being understood, is simply having the expressed attitude recognized. It does not require the hearer to respond in any further way. The hearer need not even attribute a belief, desire, or some other attitude to the speaker. Identifying the attitude the speaker is expressing, whether or not one actually attributes it to her, is just the sort of thing that can be done by way of recognizing the speaker's intention for one to do it.

Reflexive Paradox?

Commenting on the notion of an intention "to produce an effect in an audience by means of the recognition of this intention," Grice remarked, "This seems to involve a reflexive paradox, but it does not really do so" (1957/1989: 219). It seems to because the intention is self-referential. Indeed, the air of paradox may seem to imbue the hearer's inference, inasmuch as the hearer is to identify the speaker's intention partly on the

supposition that he is intended to do so. If meaning is supposed to be what people do whenever they speak to one another, there had better be nothing paradoxical about it.

It might seem paradoxical if one confuses iterative intentions with reflexive ones, as indeed Grice himself seems to have done. As Harman explains,

> Grice himself originally states his analysis as involving a self-referential intention . . . but, because of worries about what he calls "self-reflective paradox," he goes on to restate the analysis as involving a series of intentions, each about the preceding one. This turns out to lead to tremendous complexity in the resulting theory. Much of this complexity is artificial and due to Grice's refusal to stick with the original analysis and its appeal to a self-referential intention.
>
> (Harman 1986: 87–88)

As Harman stresses, Grice's move to iterative intentions led to increasingly complex formulations beginning with Strawson's (1964), followed by Grice's (1969) own, and culminating with Schiffer's (1972), each prompted by counterexamples to the previous formulation. As Harman suggests, sticking with a self-referential intention rather than invoking iterative ones avoids this complexity.

Resistance to reflexive intentions has been based not just on fear of paradox but also, it seems, on a misconstrual of what goes into the content of a communicative intention and into the hearer's inference to it (Bach 1987, in reply to Recanati 1986). Here is the simplest case of this. It might seem that Grice's formulation, with its key phrase "by means of the recognition of this intention," requires that to understand the speaker the hearer must engage in some sort of circular reasoning. It sounds as though the hearer must already know what the speaker's communicative intention is in order to recognize it. However, this misconstrues what the hearer has to take into account in order to recognize the speaker's intention. The hearer does not infer that the speaker means that p (or is expressing, say, the belief that p) from the premise that the speaker intends to convey that p. Rather, operating on the assumption that the speaker, or any speaker, in uttering a sentence intends to communicate something or other (Bach and Harnish (1979: 12) call this the "Communicative Presumption"), the hearer takes this general fact, not the identity of the specific intention, into account in identifying that intention. The situation is analogous to that of the player in those coordination games we discussed. In both cases one has to figure out what one is intended to figure out partly on the basis that one is intended to, but not by knowing in advance what it is.

Saying One Thing and Meaning Something Else

Grice elaborated on the case in which a speaker says one thing and means something else instead or something else as well. Although he did not expressly invoke Austin's notion of locutionary act, he did rely on a notion of saying whereby what is said is "closely related to the conventional meaning of the . . . sentence . . . uttered" (1975/1989: 25). However, it is not identical to conventional meaning because there can be ambiguity or context-dependent reference. Usually only one conventional (linguistic) meaning is operative in a given utterance, and linguistic meaning does not determine what such words as 'she', 'this', and 'now' are used to refer to. But even with all that fixed, the speaker might not mean just what she says.

Implicating

Grice coined the term *implicature* for what a speaker means but does not say. Whereas what is said may *imply* something, what a speaker *implicates* is a matter of her intention in saying it. For example, suppose we're dining at a greasy spoon and I ask you what you think of the pasta you're eating. You reply with (11),

(11) It's covered with Parmesan.

probably implicating that the pasta is not very good. That's a matter of your communicative intention. But how do I recognize your intention? Grice's answer would be roughly this. Presumably you are giving an informative answer to my question, but saying the pasta is covered with Parmesan is on the face of it not very informative. So you must intend me to infer that you mean something more informative. Since I asked you for your assessment of the pasta and saying that it is covered with Parmesan, good though that is, does not tell me much about the pasta itself, you are inviting me to read an assessment of it into the fact that you said what you said and no more. I can infer that you have nothing else good to say about it, hence that you mean that it is not very good. I make such an inference partly on the basis that you intend me to. Notice that what I infer is that you *mean* that it is not very good. To understand you I do not have to infer that it is not very good or even that you believe that, contrary to Grice (1975/1989: 31).

Uttering something like (11) is not the only way that you could have conveyed without saying that the pasta was not very good. Suppose you had uttered (12), in a sarcastic tone:

(12) That's the best pasta I've ever tasted.

Here you intend me, taking it to be obvious that you don't mean that it's the best pasta you've ever tasted, to infer that you mean that the pasta is not very good. You intend the fact that you said something relevant but obviously false to be my basis for figuring out what you mean.

In explaining what goes on in such cases, Grice proposed a general "Cooperative Principle" and specific "maxims"—of *quality*, *quantity*, *relevance*, and *manner*—to account for how speakers manage to make themselves understood even when they do not spell out what they mean. Actually, these maxims are better thought of as presumptions, for it is on the presumption that the speaker is being truthful, relevantly informative, and otherwise appropriate that the hearer figures out what the speaker means. By saying something obviously false (or unjustified), underinformative (or perhaps overinformative), irrelevant, or in a strange way (say by being longwinded or pedantic), one calls the hearer's attention to something ostensibly inappropriate about the utterance and, in effect, invites the hearer to reinterpret it, perhaps by considering what else one could have said instead, in such a way that it is appropriate after all. Note, however and contrary to popular opinion (see Bach 2006), that the maxims (or presumptions) do not apply only to implicature. Even when a speaker is being completely literal, meaning what she says and nothing else, the hearer reasons in essentially the same sort of way, although the reasoning is simpler, since he doesn't have to reinterpret the utterance. Also, the Cooperative Principle should be not understood to mean that interlocutors are or ought to be generally cooperative. It specifically concerns the aim of communication, not the further perlocutionary aims that people have in saying things to one another.

Implicating and Indirect Speech Acts

Implicating is a kind of indirect speech act, a special case of performing one illocutionary act by way of performing another. Consider our earlier example (5) and as well as (13), both uttered by you at a party of yours.

(5) There's some beer in the fridge.
(13) The gendarmes are coming.

In uttering (5) you could not only *tell* someone where the beer is but also indirectly offer them some. With (13) you could not only *inform* your guests that the gendarmes are coming but also indirectly warn them to quiet down. Notice that the direct illocutionary act need not be a statement. You might directly *ask* a question with (14),

(14) Do you know it's after midnight?

to *inform* someone indirectly that it is after midnight and perhaps also to suggest that it is time to go home. In these cases you have two communicative intentions (in uttering (14) perhaps three), one corresponding to each illocutionary act, and you intend your audience to recognize one by way of recognizing the other.

Between Saying and Implicating

In contrasting saying and implicating, Grice allowed both for cases in which the speaker means what he says and something else as well (implicating and indirect speech acts generally) and ones in which the speaker says one thing and means something else instead (nonliteral utterances, which Grice also counted as cases of implicating). We also need to allow for the case in which the speaker says something and doesn't mean anything (Bach 2001). Now Grice seems to have assumed that saying and implicating exhaust the cases of speaker meaning. He overlooked the possibility of an intermediate phenomenon, albeit one that also exploits the maxims. However, as others have since observed (Sperber and Wilson 1986; Bach 1994; Carston 2002; Recanati 2004), there are many sentences whose standard uses are not strictly determined by their meanings (even with ambiguities resolved and references fixed) but are not implicatures or figurative uses either. For example, if your child comes crying to you with a minor cut and you assure him,

(15) You're not going to die.

you do not mean that he will never die (that is false but irrelevant) but merely that he won't die from that cut. And if someone proposes going out to dinner and you say,

(16) [Sorry, but] I've already eaten.

you do not mean that you have eaten at some previous time (that is obviously true but irrelevant) but more specifically that you've had dinner that evening. In both cases you do not mean precisely what you are saying but something more specific. On the other hand, what you do mean isn't an implicature either. You are using each of the words in

(15) and (16) literally, but you are not using the sentences themselves literally, in that you are leaving part of what you mean implicit.

In other cases, what the speaker says is not merely not as specific as what she means but is incomplete in the sense of not being fully propositional. If your spouse is honking the horn at you and you shout back with (17),

(17) I'm not ready.

you mean that you are not ready to leave. And if your spouse yells back,

(18) We'll be late.

she means that you will be late for the event you are both planning to attend. In neither case is there anything in the sentence that corresponds to the implicit reference. Although the speaker means something definite, the sentences themselves, even with the references of the indexicals 'I' and 'we' fixed, lack determinate truth conditions. As with (15) and (16), though for a different reason (propositional incompleteness), what the speaker means is more specific than what the sentence means. These examples appear to violate the grammar school dictum that a sentence, unlike a mere word, phrase, or "sentence fragment," always expresses a "complete thought" (I say 'appear' because some would argue that such sentences contain hidden constituents). We might say that whereas what a user of (15) or (16) means is an *expansion* of the sentence meaning, what a user of (17) or (18) means is a *completion* of the sentence's content (this is Bach's (1994) terminology, which differs from both Sperber and Wilson's (1986) and Recanati's (2004), whose frameworks are somewhat less Gricean).

Now several of Grice's critics have pointed out that expansions and completions are not related closely enough to conventional meaning to fall under Grice's notion of what is said but are too closely related to count as implicatures. That is because what the speaker means is built directly from what the sentence means. Recanati (2004) suggests that the notion of what is said should be extended to cover such cases (he goes so far as to offer a series of progressively more liberal notions of saying), but clearly he is going beyond Grice's conception of what is said as corresponding to the constituents of the sentence and their syntactic arrangement. Sperber and Wilson (1986: 182) coined the word "explicature" for this in-between category, since part of what is meant explicates what is said. I propose calling these cases of *impliciture* (Bach 1994), since part of what is meant is communicated not explicitly but implicitly, by way of expansion or completion.

Summing Up

We have contrasted speaker meaning with linguistic meaning and examined Grice's ingenious conception of speaker meaning, or communication, as involving a distinctive sort of self-referential intention directed toward an audience. To make this conception compelling, we needed to distinguish the specifically communicative "effect" of understanding from other effects on the audience. This required bringing in broader notions from speech act theory. With this framework in place, we then considered a variety of ways speakers can make themselves understood even if when, as is often, what they mean is not what they say.

Related Topics

1.2 Semantics and Pragmatics
1.5 Implicature
1.6 Pragmatic Enrichment
1.10 Context Sensitivity.

References

Austin, J. L. (1962) *How to Do Things with Words*, Oxford: Oxford University Press.
Bach, Kent (1987) "On Communicative Intentions: A Reply to Recanati," *Mind & Language* 2: 141–154.
—— (1994) "Conversational Impliciture," *Mind & Language* 9: 124–162.
—— (2001) "You Don't Say?" *Synthese* 128: 15–44.
—— (2005) "Context *ex Machina*," in Z. Szabó (ed.), *Semantics versus Pragmatics*, Oxford: Oxford University Press.
—— (2006) "The Top 10 Misconceptions about Implicature," in B. Birner and G. Ward (eds.), *Drawing the Boundaries of Meaning*, Amsterdam: John Benjamins.
—— and Robert M. Harnish (1979), *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press.
Carston, R. (2002) *Thoughts and Utterances*, Oxford: Blackwell.
Grice, H. P. (1957) "Meaning," *The Philosophical Review* 66: 377–388; reprinted as chapter 14 of Grice 1989.
—— (1968) "Utterer's Meaning, Sentence-meaning, and Word Meaning," *Foundations of Language* 4: 225–242; reprinted as chapter 6 of Grice 1989.
—— (1969) "Utterer's Meaning and Intentions," *The Philosophical Review* 78: 147–177; reprinted as chapter 5 of Grice 1989.
—— (1975), "Logic and Conversation," in P. Cole and J. Morgan (eds.), *Syntax and Semantics*, *Vol. 3, Speech Acts*, New York: Academic Press; reprinted as chapter 2 of Grice 1989.
—— (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.
Harman, G. (1986) *Change in View*, Cambridge, MA: MIT Press.
Recanati, F. (1986) "On Defining Communicative Intentions," *Mind & Language* 1: 213–242.
—— (2004) *Literal Meaning*, Cambridge, UK: Cambridge University Press.
Schelling, T. (1960) *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
Schiffer, S. (1972) *Meaning*, Oxford: Oxford University Press.
Searle, J. (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, Cambridge University Press.
Siebel, M. (2003) "Illocutionary Acts and Attitude Expression," *Linguistics and Philosophy* 26: 351–366.
Sperber, D. and D. Wilson (1986) *Relevance*, Cambridge, MA.: Harvard University Press.
Strawson, P. F. (1964) "Intention and Convention in Speech Acts," *The Philosophical Review* 73: 439–460.

1.8

COMPOSITIONALITY

Josh Dever

Consider a token linguistic utterance, such as:

(1) A screaming comes across the sky.

(1) expresses some proposition—one we could perhaps identify as the proposition that a screaming comes across the sky but which we'll for convenience just call Tyrone. In virtue of what does (1) express Tyrone? Answers will vary depending on one's metasemantic views, but al-most everyone will agree that the underlying stratum of facts determining the expression relation is large and complicated. The relevant class of facts may involve:

1. Psychological facts: facts about intricately structured communicative intentions in me, and perhaps mental states too of you, and even of others not party to this communicative exchange.
2. Usage facts: perhaps up to the full history of how these words, and these syntactic structures, have been used throughout the history of the English language.
3. Contextual facts: when and where (1) was produced and received, what objects were salient at those times and locations, and what background cultural and worldly information is common ground for us.
4. Causal and environmental facts: the ways in which we and our linguistic expressions are causally embedded in the world around us (whether, for example, we are on Earth or Twin Earth).

For you to understand (1), you must know that it expresses Tyrone. But if Tyrone's identity depends on facts scattered across space and time, and often exceedingly difficult to ascertain, how can you have any hope of understanding (1), or *any* bit of language? The problem sounds intimidating, but at the same time we all know that the core of the solution is clear. One understands (1) by understanding the individual words that compose (1), and knowing how the meanings of individual words determine the meanings of complex expressions composed of those words.

The thought, then, is that the complexity of the full determination basis for the expressive features of (1) is *screened off* at the lexical level. The complex streams of determining information sort themselves out at the lexical level—from word down to world, things are an unwieldy mess, but from word up to sentence, things are relatively simple and epistemically tractable.

Why might we expect natural languages to have this screening structure? General considerations of the supervenience of the features of wholes on the features of their parts do not suffice here: we are asking for determination not by *all* properties of the parts but only by the specifically semantic properties of the parts; and we are asking for determination of the meanings of the complex expressions, and meanings are extrinsic features of expressions, and extrinsic features are typically not determined by features of parts. One reason to expect the screening structure would derive from epistemic/learnability considerations. Competent language users need to know what propositions are expressed by particular utterances. But language users cannot plausibly be expected to know the full determination basis for the expression relation, or to be able to calculate the proposition expressed even if they did know the full determination basis. Given the screening structure, however, it becomes unnecessary to know the full determination basis—all that is needed is to know the meanings of the individual lexical items, and the rules for determining meanings of complex expressions from the meanings of their parts. Of course, these epistemic/learnability constraints neither *mandate* nor are *automatically satisfied* by the screening structure. The complexities of the full determination basis could be screened elsewhere than at the lexical level (for example, at the level of some tractable characterization of utterance context), and the epistemic/learnability considerations would be equally answered. And screening could occur at the lexical level without answering those considerations—if, for example, the computation of meanings of complex expressions from the meanings of their component parts were enormously complex.

We might also expect the screening structure because we think that the complexities of the determination basis were already screened at the level of (for example) objects, properties, and relations, and then think that the lexical meanings simply reflect that prior screening. Perhaps the complexities of usage and individual psychology are all manifestations of the various complicated ways in which we, individually and collectively, relate to distal objects, but those complexities play no role in differentially determining expression of propositions, because the propositions are already determined by the distal objects. If the proposition expression relation is suitably closely tied to information flow from the world, and thus suitably independent of the vagaries of our psychologies, again the screening structure might be predicted.

1 What is Compositionality?

There are many forms a screening structure might take, but the one that has been of central interest in the philosophy of language is the property of *compositionality*. A meaning theory for a language is compositional just in case for any complex expression in the language, the meaning of that complex expression is fully determined by the meanings of its component expressions plus the syntactic mode of organization of those component expressions. Compositionality is thus a property that constrains how three things come together: the meanings of complex expressions, the meanings of components of complex expressions, and the syntactic mode of organization of the component expressions in the complex expression. Equivalently, compositionality is a property that imposes two constraints on the determination basis for the meanings of complex expressions. The factors that contribute to the meaning of a complex expression must be *syntactic and semantic*. Thus the meaning of a complex expression cannot be determined by the phonology or morphology of the words that make it up, or by the context in which it

is uttered except, in both cases, insofar as these factors themselves contribute to the meanings of the component expressions. (For more careful discussion of the interaction between compositionality and context sensitivity, see Pagin and Pelletier 2007, Recanati 2011, and Westerstahl 2011.) And the factors that contribute to the meaning of a complex expression must be *local*. The meanings of the words that compose the complex expression can contribute to the meaning of the complex; the meanings of other words (perhaps those in larger expressions of which the complex expression is itself a component part) cannot so contribute.

The compositionality constraint can be given equivalent *functional* and *substitutional* formulations. Functionally, compositionality is a property had by a meaning theory for a language just in case, for any complex expression in the language, its meaning is some function of the meanings of its component parts and their mode of syntactic organization. What does it mean to treat the mode of syntactic organization as an argument to a function, as required by this formulation of compositionality? We can think about this in one of two ways, depending on how we think of the constituents of a complex expression. Consider a complex expression such as "the hungry tiger," and assume that its syntactic structure is [$_{DP}$ [$_{D}$ the] [$_{NP}$[$_{A}$ hungry] [$_{N}$ tiger]]]. The *immediate* constituents of the expression are then the determiner "the" and the noun phrase "hungry tiger," while the *ultimate constituents* are "the," the adjective "hungry," and the noun "tiger." If we impose the constraint called *strong compositionality* (Larson and Segal 1995: 78–79), we require that the meaning of a complex expression be a function of the meanings of its immediate constituents and their mode of syntactic organization. Since the constituents are immediate, the syntactic organization can be specified by identifying the particular syntactic rule that allows those immediate constituents to be joined into the complex expression. In the case of "the hungry tiger," the rule is perhaps something like:

- $DP \rightarrow D + NP$

The arguments for the meaning composition function are then the semantic values of the immediate constituents, along with the syntactic rule by which they are combined. Strong compositionality thus encourages a picture on which each syntactic rule is associated with a semantic rule—the rule that determines how constituent meanings are to be combined when the constituents are syntactically combined by the given syntactic rule. This is typically called the *rule-to-rule* conception of compositionality, and it is most paradigmatically associated with the semantic programme of Richard Montague (Montague 1970, 1973).

If we allow the meaning of a complex expression to be determined by the meanings of its ultimate, rather than its immediate, constituents, we impose the constraint of *weak compositionality*. In this case, the syntactic structure cannot be treated as a single rule, and instead needs to be taken as a partial ordering of rules, or as a sequence of stages of syntactic derivation. If a meaning theory is strongly compositional, it is also, as the names would suggest, weakly compositional. Given strong compositionality, we can from the meanings of the ultimate constituents work out the meanings of the slightly-less-ultimate constituents, and then the meanings of the even-less-ultimate constituents, and on until we determine the meanings of the immediate constituents, and then the meaning of the complex. Thus the meanings of the ultimate constituents do, given strong compositionality, determine the meaning of the whole. But the reverse is not true—weak compositionality does not entail strong compositionality. If, for example,

the meaning of a sentence is its truth value, then strong compositionality requires that all sentential operators be truth-functional, while weak compositionality does not.

The *substitutional* formulation of compositionality is often easier to work with than the functional formulation. If the meaning of a complex expression is a function of the meanings of its component expressions (plus syntax), then it can be shown that expressions (*any* constituents, if we begin with strong compositionality; *only ultimate* constituents, if we begin with weak compositionality) with the same meaning can be intersubstituted in a complex expression without altering the meaning of the whole. Let ⟦.⟧ be some meaning theory. If ⟦the doctor's nephew⟧ is a function of ⟦the⟧, ⟦doctor⟧, ⟦'s⟧, and ⟦nephew⟧, then there is some function *f* such that ⟦the doctor's nephew⟧ = *f* (⟦the⟧, ⟦doctor⟧, ⟦'s⟧, ⟦nephew⟧). But if ⟦doctor⟧ = ⟦physician⟧, then *f* (⟦the⟧, ⟦doctor⟧, ⟦'s⟧, ⟦nephew⟧) = *f* (⟦the⟧, ⟦physician⟧, ⟦'s⟧, ⟦nephew⟧). So ⟦the doctor's nephew⟧ = ⟦the physician's nephew⟧. Functional compositionality thus entails substitutional compositionality; the reverse entailment holds as well. (See Hodges 2001 for more detailed proofs.)

Determining what function combines meanings of constituents to yield meanings of complexes can be a challenging task, one that requires production of a substantial component of a semantic theory. But checking for intersubstitutability of synonyms requires no difficult semantic theorizing, making the substitutional test typically easier to deploy. The substitutional test also makes it possible sensibly to raise the question of compositionality for semantic theories that do not assign reified semantic values but do have a notion of sameness of meaning (such as, perhaps, Davidsonian truth theories, which can be taken to assign to two expressions the same meaning if the two expressions are governed by axioms the same modulo replacement of mention of the one expression by mention of the other). It is an immediate consequence of the substitutional test that a language with no synonyms is trivially compositional, since there are no substitution instances that could possibly fail to preserve meaning.

2 A First Example

At a first glance, compositionality can easily seem a trivial constraint. How could the meaning of a complex expression *not* be determined by the meanings of its component parts? But even a cursory look at the linguistic data shows that natural languages, even if they are in the end compositional, certainly display features that *resist* a compositional treatment. Consider so-called *Frege cases*, such as:

(2) Lex Luthor believes Superman is vulnerable to kryptonite.
(3) Lex Luthor believes Clark Kent is vulnerable to kryptonite.

Initial appearances suggest the following:

1. (2) and (3) differ in truth value, and hence also differ in meaning.
2. "Superman" and "Clark Kent" are co-referential, and hence have the same meaning.
3. (2) and (3) have the same syntactic form, and differ only in the substitution of "Superman" in (2) for "Clark Kent" to form (3).

But these initial appearances then contradict the principle of compositionality. The substitution of synonymous expressions, while holding the syntax fixed, has altered meaning.

Is this a proof that English and other natural languages are not compositional? No—there are multiple ways to respond to the *prima facie* noncompositionality. The responses fall into four categories: three corresponding to established philosophical positions on Frege cases, and one not.

Fregeanism

Fregean semantic theories reject the claim that "Superman" and "Clark Kent" have the same meaning, by rejecting the move from coreferentiality to synonymy. Classical Fregean approaches run a two-tracked semantics, in which expressions are assigned both a semantic value of *reference* and a semantic value of *sense*. When a meaning theory assigns multiple types of semantic values, the question of compositionality can be raised for each type. Standard Fregean approaches strive for compositionality both in reference and in sense.

However, even given the Fregean doctrine that sense determines reference, compositionality at the two levels is independent. A Fregean meaning theory can be compositional at the level of sense but not compositional at the level of reference—the substitution of one expression for another coreferential but not cosensual expression need not preserve reference, if all we know is that substitution of cosensual expressions preserves sense. And a Fregean meaning theory can be compositional at the level of reference, but not compositional at the level of sense. If one expression is replaced with another cosensual expression, *reference* is preserved, since the two expressions will also be coreferential and we are assuming compositionality at the level of reference. But sense need not, in general, be preserved.

Direct Reference

We can reject the claim that (2) and (3) differ in meaning. In doing so, we must either (a) deny also that they differ in truth value, or (b) reject the tacit additional assumption that meaning determines truth value. If we are considering the sentences as types, option (b) is plausible, and may be developed along the lines of the *Hidden Constituent* response considered below. If we are considering the sentences as tokens, option (b) is more challenging, and the direct reference theorist is more likely to develop the position along the lines of (a), as in Salmon (1986). But in either case, if (2) and (3) differ in meaning, the challenge to compositionality is resolved.

The direct referentialist may attempt to soften the blow of claiming that (2) and (3) have the same meaning by also claiming that the two sentences give rise (perhaps, in many/typical contexts) to different conversational implicatures. When conversational implicatures are invoked, another level of meaning has been introduced, and the question of compositionality can be raised for it. There are additional complications in this case, since the meaning level of implicature does not seem to involve assigning any meanings to subsentential components (utterances implicate, but the component words making up the utterances do not individually implicate). However, Grice's implicature test of nondetachability (Grice 1975: 39) bears certain relations to a compositionality requirement—it demands that the implicatures of a sentence be fully determined by the standard semantic values of the component expressions of the sentence, plus context. The direct referentialist of this stripe is thus committed to a rejection of the quasi-compositional feature of nondetachability.

Hidden Constituents

We can deny that (2) and (3) differ only in the substitution of "Superman" in (2) for "Clark Kent" in (3). The overt form of the two sentences makes it appear that this is the only difference, but we could hold that there are covert components of the two sentences. We might, for example, think that belief is a trinary relation between an agent, a proposition, and a mode of presentation, and then hold that while the complement clause to the verb "believes" in (2) and (3) in each case provides the propositional relatum there is an additional hidden constituent of the sentence that provides the mode of presentation, and that the semantic contribution of that hidden constituent differs between (2) and (3). We might further think that the hidden constituent is indexical, so that the difference in meaning between (2) and (3) emerges only at the level of token contextualized utterances.

For a hidden constituent approach to resolve the challenge to compositionality, the hidden semantic values must be tied to some (presumably covert) syntactic components of the two sentences. The approach of Crimmins and Perry (1989), on which the hidden constituent is *unarticulated*, and not the semantic value of any syntactic constituent of the sentence, leaves the semantics noncompositional. The lesson generalizes: In order for compositionality to be possible, contextual impact on meaning of complex expressions must be mediated by contextual impact on meaning of individual words. Since a single word has no component parts, its meaning can vary from context to context without threatening compositionality (compositionality requires only that, trivially, the meaning of each individual word be a function of its own meaning). But if there are contextual impacts on complex expressions not mediated by contexutal impacts on component words, then two tokens of the same complex, deployed in different contexts, can have different meanings, despite both consisting of parts with the same meanings. Context-sensitivity is thus a special case of the general screening-off phenomenon discussed above: Meanings can and do depend on many disparate facts, but we ideally expect those disparate facts to be screened off at the level of lexical meaning.

Devious Syntax

The argument for non-compositionality is defeated if (2) and (3) have different syntactic forms. However, there does not seem to be any good reason in this example to expect a difference in syntactic form. Some related compositionality issues, though, are subject to plausible "devious syntax" approaches. For example, it is plausible that "that dingos will eat my baby" and "the proposition that dingos will eat my baby" have the same meaning: Both refer (in context) to the proposition that dingos will eat my baby. However, the pair of sentences:

(4) I fear that dingos will eat my baby.
(5) I fear the proposition that dingos will eat my baby.

can clearly differ in truth value—one attributes a perhaps rational worry about wildlife, while the other exhibits a strange phobia of abstracta. (See King 2002 for more detailed development of this problem.) Is there a problem here for compositionality? Only on the assumption that (4) and (5) have the same syntactic structure. But there is clearly reason to think that they do not—in (4), the verb phrase is the result of combining

a verb with a complementizer phrase, while in (5), the verb phrase is the result of combining a verb with a noun phrase. Given the rule-to-rule picture sketched above, we can have different semantic rules associated with these two syntactic combination rules, and compositionally account for the difference in meaning between (4) and (5).

3 The Strength of Compositionality

An important lesson can be extracted from the above discussion: although there are many phenomena in natural language that at first blush display non-compositional features, there are also a variety of methods for modifying a theory of language to allow compositional treatments of those phenomena. Because compositionality requires coordination among the three features of meanings of complexes, meanings of components, and syntactic structure of complexes, these methods can be sorted into three categories. Faced with an apparent violation of compositionality, we can revise our views on the meanings of the complex expressions. This is perhaps the most difficult of the three approaches to pursue, since we typically have more direct access to the meanings of complex expressions—especially whole sentences—than we do to the meanings of individual words. This approach thus faces the challenge of making a convincing case that appearances about sentence meaning are, in the relevant instances, misleading. However, given enough flexibility in sentence meanings, *an*y phenomenon can be given a compositional analysis. Some views on sentence meaning will automatically provide the requisite flexibility. For example, Horwich (1998: 155) holds a view of sentence meaning such that "once one has worked out how a certain sentence is constructed from primitive elements, and provided one knows the meanings of those elements, then, automatically and without further ado, one qualifies as understanding the sentence." On such a view, compositionality would appear to be automatic.

Apparent violations of compositionality can also be approached by adjusting the assignment of meanings to individual words. Words can be given meaning that are, in effect, deferred instructions for how to combine with a variety of possible embedding contexts. Suppose we want "the temperature" to behave differently when it combines on the one hand with "is 42 degrees" and on the other hand with "is rising rapidly" (picking out, roughly, a number in the first case and a function from times to numbers in the second case). Compositionality seems to forbid this, since the meaning of "the temperature" needs to be determined just by the meanings of its component parts, and not by facts about where (the particular token of) it is embedded. But if the meaning of "the temperature" is a function that maps the meaning of "is 42 degrees" to the true just in case the current heat level is some specific amount, and maps the meaning of "is rising rapidly" to the true just in case the function from times to heat levels has a certain local derivative, then "the temperature" will, in a compositional way, display the desired behavior when combined with either predicate.

The general tactic deployed here is that of *type lifting*, and in limited forms it plays a common role in semantic theories. For example, since Montague 1973, it has been common to take the meaning of a proper name to be, rather than an object, the second-order property of being a property possessed by that object, in order to make semantic combination with quantified noun phrases proceed more smoothly. When pursued vigorously enough, type lifting can in effect allow any given expression's semantic interpretation to be arbitrarily delayed, allowing its meaning, in effect, to depend on anything

in the sentence in which it occurs. In fact, if a sentence contains a single word whose meaning we can set in whatever way we please, we can then always construct a compositional theory that assigns to both the whole sentence and to the other component words of the sentence the pretheoretically desirable meanings. If compositionality is to be a nontrivial constraint, then, there must be substantive constraints on what meanings can be assigned to words.

Finally, apparent violations of compositionality can be approached by adjusting the syntactic form of the sentence. If we are sufficiently liberal with the syntactic theory, then any combination of sentence meanings and word meanings can be fitted together in a compositional theory (see Janssen 1997 for details). Liberality of this level may result in a syntactic structure for the sentence that bears no resemblance to what standard syntactic theory offers us (compositionality can easily be achieved, for example, if we endorse a syntactic theory according to which none of the words that seem to appear in the sentence are in fact syntactic constituents of it). But there are more standard syntactic techniques that are often appealed to in the quest for a compositional meaning theory. We may posit hidden syntactic constituents of the sentence, such as traces. These hidden constituents can then be carriers for semantic values needed to achieve compositionality. Or we may posit movement, and run the semantic analysis on a syntactic level at which certain constituents are displaced from their surface position, allowing them to interact with different parts of the sentence.

Compositionality, then, is by itself a constraint only as strong as one's commitment to particular triads of meanings of complexes, meanings of constituents, and syntactic structures. If we have firm commitments on all three of these matters, we may find ourselves unable to give a compositional meaning theory for the language. If we are sufficiently flexible on any of the three, compositionality can always be achieved. It is thus tempting to think that compositionality *simpliciter* is not the theoretically central concept but rather some strengthening of the core principle that combines compositionality with additional constraints. Here are three representative examples: First, the project of *direct compositionality*, for example, seeks to combine compositionality with a minimization of deviousness in the syntax—ideally, developing a version of the rule-to-rule approach, in which semantic analysis is run on surface syntactic forms, and each syntactic rule used in the construction of those surface forms is coupled with some specific method of combining component semantic values. (See Jacobson 2002, and numerous papers in Barker and Jacobson 2007.) Second, the project of *systematic semantics* requires that meanings of complex expressions be not just *functions* of meanings of their parts, but particularly tractable functions (see Zadrozny 1995). So, for example, a systematic semantics might require that meanings always combine by functional application, or by concatenation. Systematic semantics often seeks to explain the productivity of our linguistic understanding—why is it that if someone understands the expression "red apple" and the expression "large table", they also understand "large apple" and "red table." Compositionality alone is not sufficient to explain these facts about understanding—there could be a function that maps ⟦red⟧ and ⟦apple⟧ to ⟦red apple⟧, and also maps ⟦red⟧ and ⟦table⟧ to ⟦red table⟧ but which maps the latter two arguments via a vastly more complicated computational algorithm than the former two arguments, so that understanding of how the first two are mapped to their output fails to extend to understanding of how the latter two are mapped to theirs. But if we insist that the compositional function be sufficiently tractable, then understanding of the one case will extend reliably to understanding of the other. Third, we could pursue a semantic theory

in which the mechanism of type shifting is minimized, or even outlawed, so as to avoid the sort of easy compositionality via devious lexical meanings described above.

4 A Second Example

We close by working out in some detail a specific example of how the pursuit of compositionality can shape a meaning theory. Consider the problem of the meaning of *quantified noun phrases*, of the form 'DET NP', such as "the logician" or "most philosophers." Suppose we have already decided to work in a semantic framework that takes on board four commitments:

1. The meaning of a sentence is its truth value.
2. The meaning of a proper name is the object to which it refers.
3. The syntax of the language is binary branching, so that each complex expression has two immediate constituents.
4. The meanings of complex expressions are always determined by functional application: The meaning of one immediate constituent takes the meaning of the other immediate constituent as argument, and produces the meaning of the complex expression as output.

Consider a simple sentence, such as:

(6) Socrates laughs.

⟦Socrates laughs⟧ is a truth value, and it must be determined either by functional application of ⟦Socrates⟧ to ⟦laughs⟧, or by functional application of ⟦laughs⟧ to ⟦Socrates⟧. But ⟦Socrates⟧ is an object, not a function, so the first option is ruled out. Thus ⟦laughs⟧ must be a function from objects to truth values. For convenience, we can talk of the *semantic type* of any expression.

Using *t* to pick out the set of truth values, and *e* to pick out the set of objects, we say that sentences are of type *t* (the meaning of any sentence is an element of *t*) and proper names are of type *e*. Given any types *a* and *b*, we can define a derived type (*a*, *b*), which is the set of all functions from elements of type *a* to elements of type *b*. "Laughs," in this notation, is of type (*e*, *t*).

Given this framework, what is the semantic type of "some logician"? Given the sentence:

(7) Some logician laughs.

we know that either ⟦laughs⟧ (⟦some logician⟧)is of type *t*, or ⟦some logician⟧ (⟦laughs⟧) is of type *t*. Given that "laughs" is of type (*e*, *t*), the former requires that "some logician" be of type *e*. But if "some logician" is of type *e*, it must refer to an object, and there seems to be no plausible way to pick a specific object for it to refer to. We will thus set aside this possibility. If ⟦some logician⟧ takes ⟦laughs⟧ as argument to produce a truth value, then it must be of type ((*e*, *t*), *t*)—roughly, a property of properties.

However, we encounter a difficulty when we have quantified noun phrases in object position. Start with a name-containing sentence with a transitive verb:

(8) Socrates sees Plato.

"Sees Plato" must be of type (e, t), and hence "sees" must be of type $(e, (e, t)$. But now replace "Plato" with "some logician":

(9) Socrates sees some logician.

If "sees" is of type $(e, (e, t))$, then it can take arguments of type e only. If "some logician" is of type $((e, t), t)$, then it can take arguments of type (e, t) only. Thus neither "sees" nor "some logician" can take the other as argument, and there is no way to assign semantic value to "sees some logician" in a way consistent with our starting assumptions.

Our semantic framework thus makes it impossible to place quantified noun phrases both in subject and in object position. The heart of the problem is this: we have semantic reason to place names and quantified noun phrases in different semantic categories. But the two types of expression have the same syntactic distribution, so, given the commitment to compositionality and, in particular, the use of functional application as the method of semantic composition, they must co-occur in a given construction type by having one serve as argument in that construction type, and the other as argument-taker. But that particular trick can be used only in one construction type—we exhaust the trick when we find a typing for quantified noun phrases in subject position, and then it is no longer available when we reencounter the same difficulty in subject position.

There are various options for responding to this difficulty. We could, for example, reject the assumption that proper names are of type e, and unify names and quantified noun phrases in type $((e, t), t)$. Then intransitive verbs could be of type $(((e, t), t), t)$, and transitive verbs of type $(((e, t), t), (((e, t), t), t))$. Here we will instead hold on our starting four commitments, and pursue instead a syntactic route. The trouble could be eliminated if we could deny that quantified noun phrases occur in both subject and object positions. One path to such a denial is to assume that our syntax includes a rule of quantifier raising, which takes quantified noun phrases and moves them to the position of adjoin to the sentence. Thus the underlying logical form of "Some logician laughs" would be [$_S$ [$_{DP}$ some logician]t [$_S$ [$_{NP}$ t][$_{VP}$ laughs]]] (where t is a trace left by movement, and is of type e). Similarly, the underlying logical form of "Socrates sees some logician" would be [$_S$ [$_{DP}$ some logician]t [$_S$ [$_{NP}$ Socrates] [$_{VP}$ [$_{TV}$ sees][$_{NP}$ t]]]]. Given these logical forms, quantified noun phrases that appear, at surface level, in either subject or object position nevertheless occupy the same position (adjoin to S) at the level of logical form. Compositionality can then be respected by assigning the same semantic type to both subject and object quantified noun phrases.

But which semantic type would that be? Given our commitment to sentences being of type t, and given the modified syntactic position of the quantified noun phrases as joining with a sentence to form a sentence, we have no choice but to assign them type (t, t). (Even if we weakened our commitment to sentences being of type t, we would face the difficulty of typing the trace so that in both subject and object position *it* could produce the desired typing for the whole sentence.) Type (t, t), however, is the set of functions from truth values to truth values, and clearly we will not find in this type the semantic values we need to assign plausible values to quantified noun phrases.

An empirically adequate and compositional semantics for quantified noun phrases can be obtained from this impasse by altering our conception of the type t. Following Tarski 1936, we can take the fundamental semantic values for sentences to be not the

monadic features of truth and falsity but rather the relational features of satisfaction by sequences. Letting *t* now be the set of values *satisfied by sequence* σ_1, *satisfied by sequence* σ_2, and so on, we discover that since there are many sequences, *t* becomes a much richer semantic type. That means that the type (*t*, *t*) contains many more functions than before, and indeed is now a domain in which we can find plausible semantic values for quantified noun phrases.

The particular route to compositionality set out here is only one option among many (see Heim and Kratzer 1998: 178–208) for much more detailed discussion), but it illustrates a general point. Careful attention to the demands of compositionality often allows the theorist of language to uncover tacit commitments to unexpected forms of semantic value. In the case of quantified noun phrases, a priority of satisfaction to truth is revealed; the Fregean will see a discovery of a commitment to senses as emerging in the same manner; the dynamic semanticist may take the compositional treatment of cross-sentential anaphora given in Groenendijk and Stokhof 1991 as another case in point in which the necessity of context-change semantic values is revealed. By requiring coordination between meanings of complexes and meanings of their component parts, compositionality provides a guard against unpaid semantic bills—commitments regarding the meanings of sentences must eventually be traced down to their consequences for the meanings of individual words.

References

Barker, C. and P. Jacobson (2007) *Direct Compositionality*, Oxford: Oxford University Press.

Crimmins, M. and J. Perry. (1989) "The Prince and the Phone Booth: Reporting Puzzling Beliefs", *Journal of Philosophy* 86: 685–711.

Grice, H.P. (1975) "Logic and Conversation", in P. Cole and J. Morgan (ed.), *Syntax and Semantics, 3: Speech Acts*, New York: Academic Press. Reprinted in H. P. Grice (eds.), *Studies in the Way of Words*, Cambridge: Harvard University Press.

Groenendijk, J. and M. Stokhof. (1991) "Dynamic Predicate Logic", *Linguistics and Philosophy* 14: 39–100.

Heim, I. and A. Kratzer (1998) *Semantics in Generative Grammar*, Oxford: Wiley-Blackwell.

Hodges, W. (2001) "Formal Features of Compositionality", *Journal of Logic, Language, and Information* 10: 7–28.

Horwich, P. (1998) *Meaning*, Oxford: Oxford University Press.

Jacobson, P. (2002) "The (Dis)Organization of the Grammar: 25 Years", *Linguistics and Philosophy*, 25: 601–26.

Janssen, T. (1997) "Compositionality", in J. van Benthem and A ter Meulen (eds.), *Handbook of Logic and Language*, Cambridge: The MIT Press.

King, J. (2002) "Designating Propositions", *Philosophical Review*, 111: 341–71.

Larson, R. and G. Segal. (1995) *Knowledge of Meaning: An Introduction to Semantic Theory*. Cambridge:, MA The MIT Press.

Montague, R. (1970) "Universal Grammar", *Theoria* 36: 373–98.

Montague, R (1973) "The Proper Treatment of Quantification in Ordinary English", in K.J.J. Hintikka, J.M.E. Moravcsik, and P. Suppes (eds.), *Approaches to Natural Language*, Dordrecht: Reidel.

Pagin, P. and J. Pelletier (2007) "Content, Context, and Composition", in G. Peter and G. Preyer (eds.), *Content and Context: Essays on Semantics and Pragmatics*, Oxford: Oxford University Press.

Recanati, F. (2011) "Compositionality, Flexibility, and Context-Dependence", in W. Hinzen, E. Machery, and M. Werning (eds.), *The Oxford Handbook of Compositionality*, Oxford: Oxford University Press.

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: The MIT Press.

Tarski, A. (1936). "The Concept of Truth in Formalized Languages", in A. Tarski, *Logic, Semantics, Metamathematics*, Indianapolis: Hackett.

Westerstahl, D. (2011) "Compositionality in Kaplan-Style Semantics", in W. Hinzen, E. Machery, and M. Werning (eds.), *The Oxford Handbook of Compositionality*, Oxford: Oxford University Press.

Zadrozny, W. (1995) "From Compositional to Systematic Semantics", *Linguistics and Philosophy* 17: 329–42.

Further Reading

Dever, J. (2006) "Compositionality", in E. Lepore and B. Smith (eds.), *The Oxford Handbook of Philosophy of Language*, Oxford: Oxford University Press.

Pagin, P. and D. Westerstahl. (2010) "Comnpositionality I: Definitions and Variants" and "Compositionality II: Arguments and Problems", *Philosophy Compass* 5: 250–264 and 265–82.

Szabo, Z. (2007) "Compositionality", in *Stanford Encyclopedia of Philosophy*. (Three recent overviews of compositionality that deal more carefully with a number of complexities that arise in the formulation and application of a principle of compositionality.)

1.9

FOCUS AND INTONATION

Daniel Büring

1 Introduction and Terminology

Intonation describes the way the fundamental frequency of the voice, also called its pitch or F°, changes over the course of an utterance. A slightly broader term is Prosody, which covers not just intonation but also additional aspects of phonetic realization such as pauses, lengthening of segments, perhaps loudness, and spectral tilt; intonation in particular, and perhaps prosody in general, roughly corresponds to the colloquial term 'inflection'. Certain aspects of prosody (and intonation) are grammatical in nature and as such represented in a phonological representation, called Prosodic Structure. At a minimum, prosodic structure will encode prosodic constituent structure, relative metrical strength or stress of syllables, and location and nature of certain tonal (or 'intonational') events (see Ladd, 1996).

Focus—at least in connection with languages such as English—is the term used to describe effects of prosody on meaning and vice versa. The basic idea is that prosodic prominence in an unusual place serves to emphasize the meaning of the element(s) thusly marked. For example *Kim* is assumed to be focused in (1b) but not (1a) or (1c) and *Harry* is focused in (1c), but not (1a/b) (we will return to the question of what if anything is focused in (1a) in section 4.1 below):

(1) a. Kim will read Harry's book.
    b. KIM will read Harry's book.
    c. Kim will read HARry's book.

To come to terms with the relation between focus and intonation, the following simplified picture may be instructive: each sentence S has a 'neutral intonation' (part of its 'neutral prosody'), its citation form, the way it is uttered out-of-the-blue. Neutral prosody is modeled by the prosodic structure grammar assigns to S (where S is a syntactic tree, not just a string of words) on purely structural grounds. Where the prosody of a sentence S does not correspond to that neutral intonation, this is the reflex of additional, meaning-related marking, the sentence's focusing, or more generally, its Information Structure (where information structure is simply a cover term for whatever semantic or pragmatic categories such as focus, background, givenness, topic etc. one thinks may influence prosody in this way).

Roughly, neutral intonation is characterized by local frequency maxima (or sometimes: minima), so-called PITCH ACCENTS (PAs), on (virtually) all content words. MARKED PROSODY most typically omits PAs on content words at the end of the sentence (e.g. *book* in (1c)).

This in turn yields the prosodic impression that the last pitch accent (which is *not* on the final content word, as in the neutral prosody) is the strongest, most prominent (e.g. *Harry* in (1c)). The semantic or pragmatic intuition is that omitting final PAs yields the effect of emphasizing the last constituent that *does* bear a pitch accent.

It is common (though perhaps bad) practice to only indicate the last PA in orthographic renderings, as in (1) above, and not indicate any PAs in the neutral pronunciation. An alternative more faithful to the actual intonation is as in (2a) and (b), corresponding to (1a) and (c) respectively:

(2) a. KIM will READ HARRY'S BOOK.
b. KIM will READ HARRY'S book.

In (2), the final PA is marked by larger capitals, reflecting speakers' judgments (mentioned above) that the last PA—even in neutral prosody, (2a)—is more prominent than the previous ones (interestingly, phonetically, the last PA is not always more elaborate—higher, longer or louder—than the others; its main distinctive feature being, it seems, its relative position in the sentence). The last PA is also referred to as the NUCLEAR PITCH ACCENT (NPA). Marked intonation can thus also be seen as a (leftwards) shifting of the NPA; its interpretation, on this view, is emphasis on the element bearing the shifted NPA.

Since Jackendoff (1972), a privative feature [F] is used to mark focused constituents in the syntactic representation. [F] yields the interpretation of emphasis (or something like it, see section 3.2 below), and a prosodic realization which—in English at least—has the NPA within the [F]-marked constituent.

Perhaps [F] is just a convenient way to talk about the abstract concept focus, as opposed to the prosodic features that realize it, and to structurally disambiguate sentences with the same prosodic structure but—by pragmatic criteria—different foci. If so, ultimately grammar could interpret prosody directly, without the mediation of [F]. The details of such an elimination are very unclear at present though, so most works assume that [F] is a *bona fide* element of grammatical—specifically: syntactic—representation.

Summarizing our terminological usage: When people hear utterances, they perceive prominence on certain elements; this perception is based on prosodic properties, most significantly the placement of pitch accents. On the meaning side, people perceive emphasis, the grammatical counterpart to which is interpretation as focus (more on which below), or, more generally, in terms of information structure. The syntactic feature [F] mediates between emphasis and prominence, via effects on interpretation and prosodic representation.

We haven't used the term STRESS in the discussion of focus so far. Much of the nonphonological literature uses 'stress' and 'accent' interchangeably, usually to describe PAing (or its perceptive counterpart, prominence). In phonology, stress more generally describes a strength relation between syllables which can be realized phonetically in a variety of ways, including accented (more stressed)—not accented (less stressed), nuclear accent (more stressed)–pre-nuclear accent (less stressed), longer (stressed)–shorter (less stressed), full vowel (more stressed)–reduced vowel (less stressed), or

perhaps sometimes not at all. Crucially, stress is a relational, and hence graded notion (though people occasionally talk about 'word stress', 'phrasal stress', 'sentential stress' as a shorthand for 'strongest stress within the word/phrase/sentence'), while accenting is absolute (a syllable is accented or not), and accordingly, a syllable may be stressed to a certain degree and still be accentless. It is plausible to think that, ultimately, focus is realized by stress, which in turn results in pitch accenting by prosody-internal principles. For the purpose of this overview, however, we will stick to the more directly observable correlate, accenting.

2 Semantic Meaning of Focus

2.1 Associated Focus

Focus may influence truth-conditional meaning, but it doesn't itself have truth-conditional meaning. If Kim brought Harry's book and comb, (3a) is false, but (3b) may be true; if Kim brought Harry's and Jo's book, (3b) is false, but (3a) may be true:

(3) a. Kim only brought Harry's BOOK to the meeting.
b. Kim only brought HARry's book to the meeting.

Leaving out *only* in both sentences makes them truth-conditionally equivalent (though of course they still convey different emphases). That is, FOCUS SENSITIVE ELEMENTS (FSEs) such as *only* change truth conditions depending on the focus they ASSOCIATE with; focus alone doesn't.

The contribution of focus to the meaning of FSEs such as *only* can be modeled using ALTERNATIVE SETS, also called FOCUS SEMANTIC VALUES (Rooth, 1985; vonStechav, 1989). Whatever type of meaning a constituent X ordinarily has, its AS, written $[[X]]_F$, is a set of meanings of that type. If X contains a focus, $[[X]]_F$ corresponds to the meanings of expression that are like X but with the focus replaced by an alternative. Thus the VP in (3a) has an alternative set of properties (i.e. meanings!) such as {brought Harry's book, brought Harry's bike, . . .} that in (3b) like {brought Joe's book, brought Frida's book, . . .}. The FSE then relates the AS of its argument to its ordinary denotation; roughly, *only VP* will denote the set of individuals that have the property denoted by VP but not other properties from VP's AS.

2.2 Free Focus Has No Truth-Conditional Content

As for FREE FOCUS (i.e. focus not associated with a FSE), it is sometimes suggested that it introduces an existential presupposition, e.g. *KIM took Harry's book* would presuppose that someone took Harry's book.This idea has been refuted in various ways. First, the sentence can answer the questions in (4), neither of which implies (and hence license a presupposition to the effect) that someone took Harry's book:

(4) a. Who, if anyone, took Harry's book?
b. Did anyone take Harry's book?

Second, a sentence like (5) would be falsely predicted to presuppose that someone took Harry's book (and hence be contradictory) (Jackendoff, 1972):

(5) NOBody took Harry's book.

Third, there is clear difference between (6a), which—due to the cleft construction—does presuppose that someone took Harry's book, and (6b)—plain free focus—which doesn't (Rooth, 1999):

(6) a. Was it KIM who took Harry's book?
b. Did KIM take Harry's book?

Similarly, it is sometimes suggested that free focus entails exhaustiveness of the answer. But again, there is notable difference between plain focus, (7a), and a cleft, (7b), with only the latter truly entailing exhaustivity (and hence creating a contradiction)(Rooth, 1999):

(7) a. (Who attended the meeting?) KIM attended the meeting. Maybe Jo did too.
b. (Who attended the meeting?) It was KIM who attended the meeting. # Maybe Jo did, too.

Data like these show that existence and exhaustivity may be implicatures of focus but not parts of truth conditional meaning. This leaves us to look for the meaning of free focus elsewhere.

3 Pragmatic Meaning of Free Focus

3.1 Mentalist and Discourse-Related Approaches

An intuitive characterization of focus meaning would be that focus marks what the speaker intends to highlight, or emphasize, or regards as most important, or most informative.This is an example of what I will call a MENTALIST theory of focus meaning. The obvious problem with the above mentalist characterization is that the attitudes it ascribes to speakers are difficult to verify independently, or even elicit systematically.

A variant of this problem regards characterizations like "focus presupposes the existence of closed set of relevant alternatives": this is either almost trivially true, if we understand 'existence' to be about actual, real-world existence, or again immensely difficult to track down, if we talk about existence in the speaker's (or addressee's) mind.

Even more formal-looking characterizations like "focus creates/presupposes an open formula and then provides a unique value for the variable in that formula" or "focus instructs the hearer to open a new a file card and write something on it" are haunted by basically the same problem: It presupposes the reality of a certain kind of representation in the participants' minds (assuming in this case that formulas wouldn't be supposed to be objects in the real world), to which independent access would be needed in order to make the claim falsifiable.

For this reason, most formal work on the interpretation of free focus has instead taken felicity in a discourse to be the main data to be accounted for: which focusing of a given sentence is felicitous as an answer to a particular question, or in response to a particular statement, or as a narrative continuation of a previous sentence. The meaning of focus

under that view consists of discourse appropriateness conditions. Call these DISCOURSE RELATED approaches to focus meaning.

(One may argue that on such a view, focus doesn't have *meaning* so much as merely a pragmatic function; this is a valid point, though I will continue to use the term meaning in what follows.)

To give an example, answers to questions show a very systematic focusing pattern, describable as in (8):

(8) Question/Answer Congruence (QAC):
In an answer, focus marks that constituent which corresponds to the *wh*-phrase in the question.

(9) Q: Who did Jones's father recommend for the job?
A: He recommended JONES for the job.

(A formal rendering of) (8) can be seen as part of a discourse grammar, providing a partial analysis of the meaning of focus via the notion "felicitous focusing in a (question-answer) discourse."

At the same time, many if not most approaches will—implicitly or explicitly—assume that, truly, QAC in (8) doesn't constrain the relation between a focused declarative sentence and a preceding interrogative sentence but between a focused declarative sentence and some sort of *context representation construed on the basis of the question* (and maybe additional context as well). This allows the analyst to extend the analysis to cases in which, say, a pertinent question is assumed to be on participants' minds but never explicitly uttered:

(10) (A and B find the door to the classroom locked. Says A:) JONES has keys to this room.

Here we can assume that the question of who has keys to the room is on A and B's mind, even though it hasn't been uttered. A plausible analysis assumes that focus in A's answer is licensed by question/answer congruence (QAC), where 'question' is understood as the QUESTION UNDER DISCUSSION, QUD, in a discourse model. Overt QAC as in example (9) above would then be the special case of a QUD set by an actual question utterance.

So even discourse related approaches to focus are mentalist in this sense. But in contradistinction to approaches that talk about 'importance', 'highlighting' or 'open formulae in speakers' minds, discourse related theories show a direct way in which to access the pertinent mental discourse representation independently: *When* there is an explicit cotext (e.g. an interrogative utterance), it has a defined effect on context representation (e.g. setting the QUD); and all relevant aspects of context representation *can* at least be manipulated by particular utterances such as explicit questions, statements, etc.

3.2 Alternative Semantics for Free Focus

Let us assume, then, that discourse context, as construed in conversationalists' minds, and to be modeled in a formal approach to focus, includes a question under discussion (which in turn may be set by an explicit utterance or through more indirect means). Question/answer congruence can then be defined in terms of AS, as introduced in

section 2.1 above, as follows. The AS of a declarative sentence will be a set of sentence meanings, for concreteness: a set of propositions. A question denotation, too, is modeled as a set of propositions, roughly the set of all literal answers (true or false). Then QAC can, to a first approximation, be stated as follows:

(11) A declarative sentence S matches the QUD only if every proposition in the question meaning is an element of the AS of S. formally: QUD $\subseteq [[S]]_F$

Another discourse context that typically determines focus is found in so-called CONTRASTIVE FOCUS:

(12) Contrastive Focus: Focus marks the constituent that distinguishes a sentence (or parts thereof) from a previously uttered one.
T: Last year Jones' father recommended Smith for the job.
C: This year he recommended JONES for the job.

The relation between T (for target) and C (for contrast) is simply that the denotation of T must be an element of the AS of C: $[[T]]_O \in [[C]]_F$ (the same rule captures corrections as a special case of contrast).

As Rooth (1992) points out, both (12) and (8)/(11) can be seen as anaphoric processes, in the sense that T and Q(UD), respectively, are salient antecedents to which the focusing in A/C relates. However, an essential disjunction remains, as the antecedent is a subset of the AS in the Q/A case but an element of AS in the contrastive case.

Finally, some more terminology: In (12C), the entire sentence is the DOMAIN OF THE FOCUS *Jones*. What that means is that we work with alternatives at the sentence level (propositions), not the alternatives to the focus alone (individuals). Syntactically, the domain consists of the focus and its BACKGROUND; semantically, the background corresponds to the invariable part of the alternatives ('Jones's father recommended . . . for the job'), the focus to the variable one (though of course, the alternatives, being set-theoretic objects, do not literally have an internal structure, i.e. there are no semantic objects that correspond to focus and background in isolation).

In some cases, the domain of the focus is smaller than the whole sentence (and accordingly the alternatives are not propositions), e.g. VP in (3), or, as in (13), DP (from Rooth, 1992, note that there is no contrast at the higher nodes, say, 'met an X farmer').

(13) A CaNADian farmer met an AMERican farmer.

Focus domain—and derivatively, background—are essential in understanding focus; the notion of focus only makes sense as *focus within a domain*. The focus is that element within its domain that makes the (denotation of the) domain different from the target (or the question). As (9) and (12) aptly illustrate, the focus itself needn't even be new (Jones is previously mentioned in the immediate context), it is only 'new' relative to its background, in its domain. Focus is thus by its very nature a relational phenomenon; it would be a hopeless task to try to grapple with the formal modeling of focus independent of its background.

3.3 Given and New

As just pointed out, a notion like 'not previously mentioned', or plainly 'new', is not sufficient to model focus. On the other hand, focus alone is not sufficient to explain (marked) accent patterns generally:

(14) (Context: We got a toy car for Veronica. What else might she like?)

a. She'd probably also like a gaRAGE for her car.
b. A gaRAGE for her car.
c. She'd probably like a bed for her DOLLs.

(15) (What did you buy Max for his birthday?)
I bought him a WATCH at the anTIQUE store.

The focus in (14a) must be the entire object *a garage for her car* (cf. also the term answer in (14b)), yet the accent falls on the head noun *garage*, not the phrase final PP, as usual (e.g. (14c)). The intuitive reason is that (*her*) *car* in (14a/b) (but not (*her*) *dolls* in (14c)) is anaphoric, or GIVEN.

Conversely, while the answer-focus in (15) must be *a watch*—which, to be sure, is accented—there is another pitch accent on the final PP *at the antique store*, ostensibly because it is not given.

These examples show that givenness is not just the inverse of focus. Given elements can occur within a focus, and then get DEACCENTED as in (14a/b), and nongiven elements can be outside of a focus and be accented, see (15). ('Deaccenting' is the usual term for this phenomenon: An element in a syntactic position which would, in default prosody, receive an accent, goes accentless. One shouldn't literally assume, though, that there once was an accent on that very element which then is removed by some kind of transformation.)

There is also a crucial difference in formal properties: Roughly, an element is given if its meaning is salient in the discourse, prototypically if some synonymous or semantically close related expression (e.g. a hyperonym) has been mentioned before. As such, givenness is privative: One can say of an expression in context, but regardless of its syntactic environment, whether it is given or not. As emphasized in section 3.2 just above, this is different for focus, which is by its very nature relational.

That being said, it is worth noting that for a constituent C to be given is a degenerate case of contrastive focusing: namely a focus domain without a focus in it. To sketch the reasoning: If C doesn't contain a focus, its AS is the singleton set containing its ordinary meaning. Thus, any 'target' for that focus domain would, in order to be an element of C's AS, have to be synonymous with C. Building on this, Schwarzschild (1999) provides a unification of contrastive focus and givenness, which he calls Givenness (with a capital G). The remarks about the difference between focus and givenness made above do not apply to Schwarzschildian Givenness (see op. cit. for further details).

4 Grammatical Properties of Focus

4.1 Size: Broad, Narrow, Sentence-Wide

Using questions and contrasts, we can pragmatically diagnose focus and hence become more clear about the relation between [F]-marking and its prosodic realization. For

example, (16) can answer any of the questions in (16a), or—preceded by 'No!'—correct any of (16b). (17) answers/corrects (17a/b). Crucially, swapping the intonations between (16) and (17) yields infelicitous discourses:

(16) Excess heat damaged the FUSilage.

a. (i) What happened?/What caused the accident?
(ii) What effects did the excess heat have?
(iii) What did the excess heat damage?
b. (i) Someone shot at the plane.
(ii) Excess heat melted the windows.
(iii) Excess heat damaged the internal wiring.

(17) Excess HEAT damaged the fusilage.

a. What damaged the Fusilage?
b. A saboteur damaged the fusilage.

Such a pattern indicates that the NPA on *fusilage* in (16) (which is where the NPA falls in neutral intonation) is compatible with focus (and hence [F]) on the entire sentence, the verb phrase, or the object nominal, whereas the 'shifted' NPA on *heat* in (17) can only signal subject focus.

Sentence (16), as a reply to (16ai)/(16bi), is called WHOLE-SENTENCE FOCUS all other cases in (16) and (17) NARROW FOCUS. Note that narrow foci can be larger than a single word, e.g. VP in the context of (16aii)/(16bii). The term BROAD FOCUS is sometimes used synonymously with whole-sentence focus, sometimes somewhat vaguely to mean foci larger than a single word or phrase.

The size of the focus is occasionally referred to as the scope of the focus, or its domain. We discourage both usages: Since foci can be scope-bearing elements (e.g. quantifiers), 'scope of a focus' to mean its size invites confusion with the logical scope of the focused element. On the other hand, we use domain of the focus to mean the larger constituent which includes the focus and its background (see section 3.2 above).

It turns out that what we referred to as the normal intonation of a sentence systematically coincides with the realization of a whole-sentence focus. From this we may conclude that either all sentences have a focus (and the apparently focus-less ones are simply whole-sentence-foci), or that focusing only changes normal prosody in case there is a contrast between focus and nonfocus within a sentence.

We can now make precise some more terminology: Sentences may have normal intonation, or marked intonation. The former may correspond to sentence-wide focus (or perhaps the absence of focus altogether), the latter to non-sentence-wide (=narrow) foci (or some other IS-related property). The inverse is not true: Some narrow focus patterns are realized by neutral intonation (e.g. object focus in (16aiii)/(16biii)). This is why we speak of marked intonation and narrow focus but not focus intonation and marked focus.

There is an independent question whether, apart from the *placement* of PAs, there aren't any other cues to the presence of (narrow) focus, prime candidates being the shape (rising, falling, sharp rising . . .) and scaling (higher than previous, lower than previous . . .) of PAs. The answer to that question is not clear, but it seems that while there are unambiguous phonetic cues to narrow focus, these are not obligatorily present in all narrow focus sentences. For example (16), rendered with four pitch accents of roughly

descending height can answer any of the questions in (16a) (or correct any of (16b)), but with only one PA, on *fuselage*, or the PA on *fuselage* significantly higher than the previous ones, may introduce a bias to understand it as a narrow focus answer to (16aiii) (or a correction to (16biii)).

In closing we note that a single sentence may contain more than one focus:

(18) a. (Who wants tea, or coffee?) KIM wants TEA.
b. (The floor was scratched when John moved the dresser.—No! It was scratched when) SUE moved the FRIDGE.

Moreover, there may be sentences with discontinuous single foci:

(19) a. (What did you do to John?) I TOLD him about the GAME.
b. (What happened to the book?) MICHELA sent it BACK to the LIBrary.

Alternative semantics for focus as sketched in section 3.2 provides a way of handling such cases as well by interpreting both foci *in situ*.

4.2 Types: Corrective, Presentational, New-Information

Apart from differences in focus size, it is regularly suggested that there are different types of focus, such as informational, contrastive, corrective, etc.

Informally speaking, informational focus, sometimes also called presentational focus or new-information focus, would generally mark the nongiven parts of an utterance (see section 3.3) without necessarily establishing a contrast to a particular alternative (in many cases, then, answer focus is, or at least can be, presentational). Contrastive focus, on the other hand, juxtaposes two, perhaps always exclusive, alternatives; the prime instance of this would be corrections.

The crucial question is what evidence we have that these are distinguished by *grammatical* means, e.g. different pitch accent patterns or types. Prosodic evidence in English seems inconclusive. In particular, it is hard to distinguish between the claim that a particular focus is, for example, corrective (and hence realized, say, with a larger pitch excursion), and the claim that an ordinary), narrow focus is realized in a contrastive utterance (which therefore has a larger pitch range in general).

On the other hand, Italian and many other Romance languages, have a focus fronting construction in which an intonationally distinguished, and pragmatically focused, element appears in noncanonical left-peripheral position. Focus fronting never happens with answer foci (these must be realized right-peripherally) but typically occurs in corrections or explicit juxtapositions. This kind of pattern can be taken as evidence that grammar does distinguish different kinds of foci, although analyses that aim to derive the differences while maintaining a unified concept of focus have been proposed (Brunetti, 2004; L'opez, 2009; Samek-Lodivici, 2006).

It is important to distinguish constructions like Romance focus fronting, which (ostensibly) marks a particular pragmatic type of focusing, from constructions like the English cleft (which is also occasionally referred to as a focusing construction). English clefts have particular *semantic* properties: namely an existence presupposition and an exhaustivity claim. As we saw above, focus conversationally implicates the same things. As such, focusing and clefting often align for pragmatic reasons as in (20a), but it is straightforward to dissociate them, (20b) (Prince, 1978: 885/898):

(20) a. It was MaGRUDer that leaked it.
b. It was ALSO during these centuries that VAST INTERNal MIGRATION from the SOUTH took place.

A more complicated case is presented by Hungarian, which has a preverbal position that has largely the same semantic properties as English clefts (existence presupposition and exhaustivity claim). Unlike with English clefts (and Romance focus fronting), however, answer foci and corrective foci virtually obligatorily appear in that position, unless they are decidedly incompatible with its semantic properties. Conversely—and also unlike English clefts, see (20b)—elements in that position are always stressed and interpreted as pragmatic foci (e.g. in Q/A sequences). É. Kiss (1998) concludes that foci in the preverbal position in Hungarian are a different type of focus (she calls them "identificational") than focus in postverbal position (É. Kiss's "informational focus"); it seems equally possible, though, that exhaustivity (and existence) are contributed by the position (like in English clefts), while the meaning of focus in pre-and postverbal position is the same (Brody, 1995; Horvath, 1986, 2010; Szendrői, 2001, 2004).

To conclude, it is hard to determine with certainty whether there are different types of grammatical focus. It is at least not obvious that intonational focus in English (and other Germanic languages) differentiates systematically between uses such as contrastive, informational, or identificational.

4.3 Focus Realization Rules

We mentioned several times above that one main principle of focus realization in English (and many other languages) is that foci bear the NPA. Truckenbrodt (1995) makes the strong claim that this is indeed the only effect focusing has on prosody, with all other effects following from prosody-internal regularities, and in particular from what may be called PROSODIC INERTIA: that those principles that determine neutral prosody remain active in focused structures as well. As a consequence, neutral prosody is retained as much as compatible with the requirement that focus bear the NPA.

Of particular importance is the assumption that the final PA in a prosodic domain (usually the clause) is grammatically the strongest. This is of course the theoretical counterpart to our earlier observation that the final PA is *perceived* as the most prominent. It means that—within a prosodic constituent—one cannot have a stronger PA followed by a weaker one (though the inverse is possible). Assuming that focus needs to contain the strongest PA, PAs after the focus must be omitted altogether (since otherwise, *they* would be strongest). Furthermore and trivially, if the focus doesn't contain a PA by virtue of unmarked prosody (say focus on a pronoun, or an auxiliary, or some other functional element, or a phrase final predicate—see below), one will be added. In a nutshell: the focus contains a PA; apart from that, everything up to and including the focus will retain unmarked prosody; everything after the focus until the end of the focus domain remains accentless.

One particularity of unmarked intonation that is retained under focusing is PREDICATE INTEGRATION, roughly: predicates (in particular: verbs) remain accentless if following an accented argument. This effect has been widely discussed in English intransitives, as well as head final verb phrases in Dutch and German:

(21) a. Your MOTher called. (*not*: Your MOTher CALLED)
b. ein BUCH lesen (*not*: ein BUCH LESen)
a book read
'to read a book'

[German]

Whatever the exact scope and characterization of predicate integration is, its is—under the view argued for here—not a consequence of some focus realization rule (as it is for Gussenhoven, 1983; Selkirk, 1984, 1995, cf. the latter's FOCUS PROJECTION RULES), but of unmarked intonation. So if we claimed above that within a focus (and before it), unmarked intonation occurs, this will include unaccented predicates in integration configurations. Such predicates will only be accented if they are focused, and their argument is not. (On integration (under various names) see among many others Schmerling (1976), Fuchs (1984, 1976), Gussenhoven (1983), Jacobs (1991/2b, 1992, 1999); and, from a slightly different vantage point, Sasse (1987), Jäger (2001); for a recent prosody-based analysis Truckenbrodt (2006).)

If, as in (21), an argument–predicate sequence happens to occur in sentence-final position, integration will lead to a 'shift' of the NPA to the left. But crucially, there is no need for a rule of NPA shift, nor indeed for a rule of NPA placement (such a rules is e.g. formulated in Cinque, 1993, and widely adopted); it simply follows from two factors: local PA omission on a predicate due to integration, and the purely prosodic definition of NPA as the final PA.

5 Summary

This article surveyed some core semantic, pragmatic, and phonological properties of focus and intonation. Perhaps surprisingly, no definition of focus was offered. Rather, I suggested a rough demarcation of phenomena potentially amenable to an analysis utilizing focus (marked accent patterns), introduced a formal tool to formulate focus-related grammatical conditions (alternative sets) and then gave a semi-formal characterization of two of the most common and best understood focus phenomena: question/answer focus and contrastive focus.

A complete theory of focus would offer similar characterizations of all phenomena within this realm, whatever they may be, and hopefully be able to unify them formally (unless, of course, they turn out to be properly grammatically distinguished in a way not attributable to other factors). At present, some generalization of contrastive focus along the general lines of anaphoric contrast seems most promising in this regard, but many details remain to be worked out.

Cross linguistic studies of focusing are of utmost importance in this project. Considering the lack of a general and sufficiently precise definition of focus, the starting point for the more successful crosslinguistic studies are again uncontroversial instantiations of the general phenomenon, namely answers and contrastive utterances. Once the grammatical realization of such constructions in a language is well understood, one can proceed to ask what other pragmatic (or grammatical) configurations in that language receive the same or a similar realization, thereby widening again the scope of focus theory in general, and allowing for meaningful crosslinguistic comparison.

Hopefully this article can serve as a guideline to researchers new to this endeavor, and aid in avoiding common terminological and methodological pitfalls.

Bibliography

Akmajian, Adrian (1973). "The Role of Focus in the Interpretation of Anaphoric Expressions." In Stephen Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*, 215–226. New York: Holt, Rinehart & Winston.

Beaver, David and Brady Clark (2008). *Sense and Sensitivity: How Focus Determines Meaning*. Cambridge: Wiley-Blackwell.

Bosch, Peter and Rob van der Sandt (1999). *Focus—Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press.

Bródy, Michael (1995). "Focus and Checking Theory." In István Kenesei, ed., *Levels and Structures, Approaches to Hungarian* 5, 31–43. Szeged: Jate.

Brunetti, Lisa (2004). *A Unification of Focus*. Padova: Unipress.

Büring, Daniel (2006). "Focus Projection and Default Prominence." In Valéria Molnár and Susanne Winkler, eds., *The Architecture of Focus*, 321–346. Berlin/New York: Mouton De Gruyter.

Chomsky, Noam (1971). "Deep Structure, Surface Structure and Semantic Interpretation." In Danny D. Steinberg and Leon A. Jakobovits, eds., *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, 183–216. Cambridge: Cambridge University Press.

Cinque, Guglielmo (1993). "A Null Theory of Phrase and Compound Stress." *Linguistic Inquiry* 24: 239–267.

Culicover, Peter W. and Louise McNally, eds. (1998). *The Limits of Syntax*, vol. 29 of *Syntax and Semantics*. New York: Academic Press.

van Deemter, Kees (1994). "What's New? A Semantic Perspective on Sentence Accent." *Journal of Semantics* 11:1–31.

van Deemter, Kees (1999). "Contrastive Stress, Contrariety, and Focus." In *Bosch and van der Sandt* (1999), 3–17.

Dekker, Paul and Herman Hendriks (1996). "Links without Locations—Information Packaging and Non-monotone Anaphora." In Paul Dekker and Martin Stokhof, eds., *Proceedings of the Tenth Amsterdam Colloquium*, 339–358.

Dik, Simon C. (1997). *The Theory of Functional Grammar. Part 1: The Structure of the Clause*. Berlin, New York: Mouton de Gruyter. Edited by Kees Hengeveld.

É. Kiss, Katalin (1998). "Identificational Focus and Information Focus." *Language* 74: 245–273.

Erteschik-Shir, Nomi (1997). *The Dynamics of Focus Structure*. No. 84 in Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

Fuchs, Anna (1984). "'Deaccenting' and 'Default' Accent." In Gibbon and Richter (1984), 134–164.

Fuchs, Anna (1976). "Normaler und kontrastiver Akzent." *Lingua* 38: 293– 312.

Geurts, Bart and Rob van der Sandt (2004). "Interpreting Focus." *Theoretical Linguistics* 30(1): 1–44.

Gibbon, D. and H. Richter (1984). *Intonation, Accent and Rhythm: Studies in Discourse Phonology*. Berlin.

Gussenhoven, Carlos (1983). "Focus, Mode, and the Nucleus." *Journal of Linguistics* 19: 377–417. Reprinted as chapter 1 in Gussenhoven (1984).

Gussenhoven, Carlos (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, Carlos (2007). "Types of Focus in English." In Chungmin Lee, Matthew Gordon, and Daniel Büring, eds., *Topic and Focus—Cross-linguistic Perspectives on Meaning and Intonation*, no. 82 in Studies in Linguistics and Philosophy, 83–100. Dordrecht: Springer.

Halliday, M.A.K. (1967). "Notes on transitivity and theme in English, part II." *Journal of Linguistics* 3: 199–244.

Hendriks, Herman (2002). "Information Packaging: From Cards to Boxes." In Kees Van Deemter and Rodger Kibble, eds., *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*. Stanford: CSLI.

Horvath, Julia (1986). *Focus in the Theory of Grammar and the Syntax of Hungarian*. Dordrecht: Foris.

Horvath, Julia (2010). "'Discourse Features', Syntactic Displacement and the Status of Contrast." *Lingua* 120(6): 1346–1369.

Jackendoff, Ray (1972). *Semantics in Generative Grammar*. Cambridge, MA: MIT Press.

Jacobs, Joachim (1991/2a). *Informationsstruktur und Grammatik*. Opladen: Westdeutscher Verlag (= *Linguistische Berichte* Sonderheft 4).

Jacobs, Joachim (1991/2b). "Neutral Stress and the Position of Heads." In Jacobs (1991/2a), 220–244.

Jacobs, Joachim (1992). Integration. Tech. Rep. 14, SFB 282, Düsseldorf, Köln & Wuppertal.

Jacobs, Joachim (1999). "Informational Autonomy." In Bosch and van der Sandt (1999), 56–81.

Jäger, Gerhard (2001). "Topic-Comment Structure and the Contrast Between Stage-Level and Individual-Level Predicates." *Journal of Semantics* 18: 83–126.

Ladd, Robert D. (1980). *The Structure of Intonational Meaning*. Bloomington: Indiana University Press.
Ladd, Robert D. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
Lambrecht, Knud (1994). *Information Structure*. Cambrige: Cambridge University Press.
López, Luis (2009). *A Derivational Syntax for Information Structure*. No. 23 in Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.
Miller, Judi (2006). "Focus." In Keith Brown, ed., *The Encyclopedia of Languages and Linguistics*, vol. 4, 511–518. Oxford: Elsevier.
Prince, Ellen F. (1978). "A Comparison of Wh-Clefts and it-Clefts in Discourse." *Language* 54(4): 883–906.
Reinhart, Tanya (2006). *Interface Strategies: Optimal and Costly Computations*. Cambridge, MA: MIT Press.
Rizzi, Luigi (1997). "The Fine Structure of the Left Periphery." In Liliane Haegeman, ed., *Elements of Grammar: Handbook of Generative Syntax*, 281–337. Dordrecht: Kluwer.
Roberts, Craige (1996). "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." In J.H. Yoon and Andreas Kathol, eds., *OSU Working Papers in Linguistics 49: Papers in Semantics*, 91–136.
Rooth, Mats (1985). Association with Focus. Ph.D. thesis, University of Massachusetts, Amherst.
Rooth, Mats (1992). "A Theory of Focus Interpretation." *Natural Language Semantics* 1: 75–116.
Rooth, Mats (1999). "Association with Focus or Association with Presupposition." In Bosch and van der Sandt (1999), 232–244.
Samek Lodivici, Vieri (2006). "When Right Dislocation Meets the Left-Periphery: A Unified Analysis of Italian Non-Final Focus." *Lingua* 116(6): 836–873.
Sasse, Hans-Jürgen (1987). "The Thetic/Categorical Distinction Revisited." *Linguistics* 25: 511–580.
Schmerling, Susan F. (1976). Aspects of English Sentence Stress. Ph.D. thesis, Austin.
Schwarzschild, Roger (1997). "Interpreting Accent." Unpublished ms. Rutgers University.
Schwarzschild, Roger (1999). "GIVENness, AvoidF and Other Constraints on the Placement of Accent." *Natural Language Semantics* 7(2): 141–177.
Selkirk, Elisabeth (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
Selkirk, Elisabeth (1995). "Sentence Prosody: Intonation, Stress, and Phrasing." In John A. Goldsmith, ed., *The Handbook of Phonological Theory*, 550–569. London: Blackwell.
von Stechow, Arnim (1989). *Focusing and Backgrounding Operators*. Tech. Rep. 6, Fachgruppe Sprachwissenschaft, Universität Konstanz.
Szendrői, Kriszta (2001). *Focus and the Syntax-Phonology Interface*. Ph.D. thesis, University College London.
Szendrői, Kriszta (2004). "Focus and the Interaction between Syntax and Pragmatics." *Lingua* 114(3): 229–254.
Truckenbrodt, Hubert (1995). Phonological Phrases: Their Relation to Syntax, Focus, and Prominence. Ph.D. thesis, MIT, published 1999 by MIT Working Papers in Linguistics.
Truckenbrodt, Hubert (2006). "Phrasal Stress." In Keith Brown, ed., *The Encyclopedia of Languages and Linguistics*, vol. 9, 572–579. Oxford: Elsevier.
Vallduví, Enric (1990). *The Informational Component*. Ph.D. thesis, University of Pennsylvania. (published 1992: Garland).
Vallduví, Enric and Maria Vilkuna (1998). "On Rheme and Contrast." In Culicover and McNally (1998), 79–180.

1.10

CONTEXT-SENSITIVITY

Tom Donaldson and Ernie Lepore

1 Introduction

We begin with a puzzle. Consider the following two utterances:

(1) I'm Spartacus! *[Said by Spartacus]*
(2) I'm Spartacus! *[Said by Antoninus]*

What Spartacus said was true, and what Antoninus said was not. Yet the two slaves uttered the exact same sentence, so how can this be?

Admittedly, the puzzle is not very hard, and its solution is uncontroversial. The first person pronoun "I" is—to use a technical term—context-sensitive. When Spartacus uses it, it refers to Spartacus; when Antoninus uses it, it refers to Antoninus. So when Spartacus says "I'm Spartacus," he expresses the true proposition that he, Spartacus, is Spartacus. And when Antoninus says it, he expresses the false proposition that he, Antoninus, is Spartacus. The sentence "I'm Spartacus" expresses different propositions when used by different people.

Another example will help. Contrast these two utterances, made by subjects in a study carried out by experimental epistemologists:

(3) This is a zebra. *[Said by someone while pointing at a zebra]*
(4) This is a zebra. *[Said by someone while pointing at a cleverly decorated mule]*

(3) is true; (4) is false—even though they are tokens of the same sentence. Why? Because the word "this," like "I," is context-sensitive. Roughly, an instance of the word "this" refers to whichever object is being indicated, or otherwise rendered salient, to the audience of the utterance. In (3), the object indicated is a zebra, and so, the token of "this" in (3) refers to that zebra. This animal is the subject of the sentence. So in saying (3), the speaker truly says, of a certain zebra, that it is a zebra. In the case of (4), on the other hand, the indicated object is a mule. In saying (4), then, the speaker says of a certain mule that it is a zebra. Utterance (4) is therefore false.

Enough examples for now—let's have some definitions. "Context," as the word is used in semantics, is a technical term. There is some dispute about exactly what we should take contexts to be, but for our purposes it suffices to say that the context of an utterance is comprised of all those features of the circumstance in which the utterance is made that are relevant to interpretation. A context-sensitive word, phrase, or sentence

is one which can receive different semantic values (interpretations) when used in different contexts. In particular, a declarative sentence is context-sensitive if it can express different propositions when said in different contexts.

We can now begin to form a list of words which are context-sensitive in an uncontroversial and straightforward way:

Pronouns: *I*, *me*, *my*, *you*, *yours*, *he*, *him*, *his* . . .
Demonstratives: *this*, *that*, *those*, *these* . . .
Adjectives: *current*, *present* . . .
Adverbs: *today*, *tomorrow*, *now*, *here*, *there* . . .

There is little dispute about the context-sensitivity of the words on this list. Most English speakers quickly acknowledge their context-sensitivity when it is pointed out to them. It is also usually thought that tense morphemes—like the ending "-ed" which marks the past tense in English—are also context-sensitive (although this has been disputed).

Other expressions, however, are the subject of controversy. Indeed, a great deal of dispute revolves around the question of how context-sensitive natural languages are. What's more, there is disagreement about how context-sensitivity should be explained in systematic linguistic theory. We will begin, in sections 2–5, by looking at the question, *Which natural language expressions are context sensitive?* We will go on, in sections 6–9, by investigating how context-sensitivity should be best accommodated within linguistic theory.

2 Further Apparent Examples of Context Sensitivity

2.1 First Example: "On the Left"

"On the left" is not as obviously context-sensitive as "I" and "this," but a little reflection suggests forcefully that it is indeed context-sensitive. Two New York tour buses pass one another on the road, going in opposite directions past the American Museum of Natural History. On one bus, the guide says, correctly:

(5) The American Museum of Natural History is on the left.

If the guide on the other bus were to say the same thing, all of the passengers would look the wrong way.

"On the left," it would seem, is context-sensitive. Roughly, in a given context, something is "on the left" just in case its direction from the contextually salient location is about 90 degrees anti-clockwise from the contextually salient "forward" direction.

2.2 Second Example: "It's Raining"

John is having a three-way phone conversation with his friends Nitika (in Bombay) and Kurt (in Vienna). The conversation goes like this:

(6) John: What's the weather like, Nitika?
Nitika: It's raining.

John: Kurt?
Kurt: It's not raining.

It seems that Nitika and Kurt could both have uttered the truth—despite the appearance of contradiction. Arguably, when Nitika says "It's raining," she says that it is raining *in Bombay*, while when Kurt says "It's not raining," he says that it is not raining *in Vienna*. Does it follow that "It's raining" is context-sensitive?

2.3 Third Example: "Heavy"

Suppose that Chiara has bought a bike as a present for her niece. She chose this particular bike because, when compared to the alternatives, it was the most lightweight. Having bought the bike, Chiara went to the post office, intending to mail the bike to her niece. She was disappointed to find that she would be charged by the kilogram, and the postage cost was enormous. Now consider:

(7) This bike is not heavy. *[Said by Chiara in the store, admiring its lightweight construction]*
(8) This bike is heavy. *[Said by Chiara in the post office, despairing at the cost of sending the bike by mail]*

Arguably, both utterances (7) and (8) are true. In the context of (7), the bike does not satisfy "heavy," but in the context of (8), it does. A natural thing to say is that—somehow—there has been a shift in the "standard" or "threshold" for heaviness.

2.4 Fourth Example: "Green Leaves"

(9) is clearly context-sensitive, because it contains the word "my":

(9) The leaves on my maple tree are green.

Charles Travis has argued that this sentence is context-sensitive in another, less obvious, way (Travis 1997). Travis asks us to consider Pia who owns a Japanese Maple tree, with russet leaves. Thinking that the leaves should be green, Pia paints them, and having finished, she says (9)—apparently truly. Shortly afterwards, Pia receives a phone call from a friend, a botanist looking for green leaves for a study of green-leaf chemistry. Pia offers him the leaves from her Japanese Maple tree. This time, when she utters (9), she says something false. The point is not that "my maple tree" has a different referent on the two occasions—the same tree is referred to both times. Moreover, the tree's leaves have not changed color between the first utterance and the second. As Travis likes to put it, it seems that what "counts as" green is different in the two contexts.

2.5 Fifth Example: Quantifiers

When someone opens his cupboard, peers inside, sighs, and reports:

(10) All the peaches are rotten.

He is unlikely to be reporting on a threat to the world fruit supply: rather, his quantifier "all the peaches" ranges just over those peaches in his cupboard or those in his kitchen, or some similar contextually salient domain. The very same sentence, used in a different context, would have different truth-conditions because the domain of quantification would be different. Arguably, then, sentences involving quantifiers are context-sensitive, because the domain of quantification is contextually determined.

3 Minimalism, and Moderate and Radical Contextualism

Multiplying examples of this sort might quickly lead one to conclude that *all* terms in natural language are sensitive to context, and in complicated ways that are hard to describe systematically. Any given declarative sentence, on this view, can express different propositions in different contexts. One might even begin to think it misleading to speak of *the meaning* of a sentence type—on the grounds that the meanings of its various tokens have so little in common. This sort of position is sometimes called 'radical contextualism'.

Radical contextualists are at one end of a spectrum of different views. At the other end of the spectrum are the *minimalists*, who deny there is context-sensitivity in natural language beyond whatever is introduced by the "easy" examples of context-sensitive expressions discussed in section [1]. Between the radicals and the minimalists there are a variety of *moderate contextualist* positions.

4 Objections to Radical Contextualism

It would be accepted by everyone that (11) is context-sensitive, because it contains the demonstrative "these":

(11) These leaves are green.

We could say that (11) has no truth-conditions *independently of context*. However, most philosophers would say that an adequate semantic theory for English would provide us with a generalization of the form:

Relative to a context c, (11) is true iff ____

that is, a generalization which tells us the truth-condition of (11) relative to an arbitrary context. One's first guess might be:

(11*) Relative to a context *c*, "These leaves are green" is true iff the objects indicated in *c* are green leaves.

This is perhaps a little crude, but it is usually thought that semanticists should seek generalizations of roughly this kind. The characteristic claim of radical contextualists is that this search will fail; motivated by examples like those discussed in section [2], radical contextualists deny that generalizations like this could capture the subtle and multifarious ways in which the truth conditions of English sentences depend on context.

This, then, is the negative claim characteristic of radical contextualism. Problems arise, however, when radical contextualists get around to defining their views positively.

In particular, it is difficult for radical contextualists to answer the question, *What does it take for an utterance of (11) to be true?* Here is Travis struggling with this question:

> What could make given words "The leaves are green" true, other than the presumed "fact that the leaves are green," is the fact that the leaves *counted* as green on the occasion of that speaking. Since what sometimes counts as green may sometimes not, there may still be something to make other words "The leaves are green" false, namely that on the occasion of *their* speaking, those leaves (at that time) did not count as green.
>
> (Travis 1997: 101–2; see, also, Travis 1996: 457)

This passage suggests an utterance of (11) can be true even when the indicated leaves aren't green, because they might still *count as* green. Apparently, as Travis sees it, the indicated leaves can count as green even if they *aren't*. This should seem strange all by itself, but the key critical point we want to register is that in saying this Travis has actually contradicted the central claim of his radical contextualism. It would seem that on Travis's view, something like the following is true:

> (11_{Travis}) Relative to a context *c*, "These leaves are green" is true iff the objects indicated in *c* count as green leaves in *c*.

But this is exactly the sort of generalization which radical contextualists claim to be impossible!

A second inconsistency charge against radical contextualism arises from Travis's answer to the question: "What, other than the indicated leaves being green, could ever suffice for the truth of and utterance of (11)?" To see what's gone wrong, note that Travis's answer itself is provided in a particular context. Let's call this 'Travis's Context', or TC for short. According to Travis, the sentences in his paper have their truth-conditions determined in TC. In particular, sentences containing the word "green" in Travis's article, must, according to radical contextualism, have the kind of context-sensitive truth conditions radical contextualism claims all sentences have. So the truth conditions for (12):

> (12) What could make the given words "The leaves are green" true, other than the presumed "fact that the leaves are green," is the fact that the leaves counted as green on the occasion of that speaking.

are determined (in part) by what counts as green in TC, i.e., the semantic contribution of the expression "green" as it occurs in an utterance of (12) in TC depends on what counts as green in TC. Let's say that anything that counts as green in TC is green_{TC}. But if this is so, (12) then should be read as (12_{TC}):

> (12_{TC}) What could make the given words "The leaves are green" true, other than the presumed "fact that the leaves are green," is the fact that the leaves counted as green_{TC} on the occasion of that speaking.

We hope it is clear this is not what Travis intends to be saying. He doesn't mean to suggest that the leaves counting as green_{TC} is what would make an utterance of "The

leaves are green" true in contexts other than TC (and (12) is clearly about contexts other than TC). The bottom line is that he seems to be trying to *use* "green" as it occurs in (12) in a context-insensitive way. Since, according to him, such context-insensitive uses are not possible, it follows that (12) either says something false (i.e., that counting as green_{TC} is what makes utterances of "the leaves are green" true in contexts other than TC) or it employs the word "green" in a way inconsistent with radical contextualism. (The same point applies to our use of "green" on the RHS of (11_{Travis}), and, for that matter, for our attempt to formulate this objection.)

In short, if these arguments are sound, then the radical contextualists have provided no coherent alternative to the standard picture. We turn now to more systematic ways of distinguishing context-sensitive expressions from those which are insensitive to context. In this next section, we will canvass a number of tests for context sensitivity: that is, procedures for identifying context-sensitive expressions.

5 Tests for Context-Sensitivity

In section [2], we explored a variety of examples which suggest that context-sensitivity in natural language extends well beyond the easy examples we discussed in section [1]. All of these arguments are more or less controversial. Moreover, taken together they push us towards radical contextualism—a position we have seen is inherently untenable. What we would like is a *test* with which systematically to distinguish expressions which are context-sensitive from those which aren't. In this section we will look at some of the tests that have been proposed for this purpose.

5.1 The Disquotational Indirect Quotation Test (IDR)

Linguistic practices indifferent to context should expose expressions that are not. We will begin by relying on (Cappelen and Lepore 2005).

Disquotation is usually a safe policy in insuring accuracy in indirect reporting. When Alice says "Bob Dole is Italian," she is usually and most straightforwardly best reported with "Alice said that Bob Dole is Italian." Context-sensitive language, however, mandates adjustments. If Bob Dole utters "I'm Italian," the indirect quotation "Bob Dole said that I am Italian" fails to report him (unless the reporter is Bob Dole). Such obvious considerations lead to a test for context-sensitivity: the easier it is to indirectly report an utterance disquotationally, regardless of indifference or ignorance about its context of use, the less likely it is that its constituents are context-sensitive. In short:

> IDR: If reporters can easily and truly indirectly disquotationally report an utterance of a sentence S by an agent A, i.e., with 'A said that S,' despite indifference about, or ignorance of, its original context of utterance, then it is unlikely S is context-sensitive.

5.2 The Collectivity Test (Collectivity)

Collective reports of utterances of a sentence often preserve truth. If John and Herman both utter "China is larger than Russia," we easily collectively report them with "John and Herman said that China is larger than Russia." An exception is when what's uttered is context-sensitive. Suppose we know John and Herman each uttered "Bill left today."

We cannot then accurately collectively report them with "John and Herman both said that Bill left today," unless we happen to know that both John and Herman spoke today. The context-sensitive word "today" blocks collection; and in general, context-sensitive expressions resist collective accurate reporting. Such obvious considerations lead to a test for context-sensitivity based on collecting what others say: The easier it is to collect distinct utterances of a single sentence into a single disquotational report, regardless of indifference about, or ignorance of, their original contexts of use, the less likely it is that the sentence uttered is context-sensitive. In short,

> Collectivity: Let u and u' be utterances of S by A and B. If reporters can easily collect u and u' into a single true indirect disquotational report, i.e., with 'A and B both said that S,' despite indifference about, or ignorance of, the original contexts of utterance, it is unlikely S is context—sensitive.

In both tests—and others, cf. (Cappelen and Lepore 2005)—it is presumed that accurate reporting requires reporter and reportees to be expressing the same thought (or proposition) when the latter's utterances easily admit of a disquotational or collective indirect report.

5.3 Agreement Tests (Agree)

When we apply the IDR and Collectivity tests to expressions like "immigrate," "left," "nearby," "it's raining," "local," "heavy," and "ready," the tests suggest that these expressions are *not* context-sensitive. For example, an utterance of (13) can often easily be reported with (14), no matter what the environment of the reporter is like, despite indifference about, or ignorance of, the original context of utterance.

(13) It's raining.
(14) John said that it's raining.

Likewise, when John and Herman each utter (13), they can be easily reported collectively with (15) across contexts, despite indifference towards, or ignorance of, John's and Herman's original contexts of utterances.

(15) John and Herman said that it is raining.

"Left" and "nearby" submit to the same collection patterns. How can an expression behave like this if it is context-sensitive? Several authors have argued that IDR and Collectivity are poor tests for identifying context-sensitive expressions. Space limitations preempt our going very far into this discussion; however, some of these authors recommend replacing indirect reporting tests with agreement tests.

Suppose John and Herman both utter (13) indicating different locations. While they may easily admit of the collective report (15), it is difficult in such circumstances to characterize them with (16):

(16) John and Herman (both) agree that it is raining.

The same applies to some other supposedly context-sensitive expressions, including those containing "nearby," "left," "local," "ready," "enough," etc. Exploiting such

judgments, some authors have recommended introducing Agree-1 and Agree-2 as tests for context- sensitivity (in place of indirect reporting tests which they see as flawed).

> *Agree-1: If A utters S, B utters its negation, and they are not easily reported as disagreeing, say, with 'A and B disagree whether S', then S is semantically context-sensitive*
>
> (Cappelen and Hawthorne 2009)

If A says "I'm happy" and B "I'm not happy," no one would report them with "A and B disagree"; this is because "I" is context-sensitive.

> Agree-2: If A and B both utter S and can be reported as agreeing, say, with 'A and B agree that S', then that is evidence S is semantically invariant across its distinct utterances. If, on the contrary, distinct utterances cannot be so reported, this is evidence S is not semantically invariant across its distinct context of utterance.
>
> (Cappelen and Hawthorne 2009: 54–55)

If both A and B utter "John bought a table," then if they can be correctly reported with "A and B agree that John bought a table," this is because "table" is context-insensitive.

In short: if speakers in distinct contexts utter S but can be reported as agreeing, then S is context-insensitive; and if one utters S while the other utters its negation, then S is context-sensitive only if they needn't disagree.

One reason for preferring agreement judgments over indirect reporting ones in ascertaining context-sensitivity is that the former do not admit of distributive readings while the latter may. Agreement and disagreement require coordination on a single proposition, while indirect reporting does not.

6 Content and Character

So far, we have been discussing the question, "Which expressions are context-sensitive?" We have yet to discuss how to explain context-sensitivity in systematic linguistic theory. It is to this topic we turn in the next few sections.

We start with the easiest sort of example:

(1) I'm Spartacus! *[Said by Spartacus]*
(2) I'm Spartacus! *[Said by Antoninus]*

As noted back in section [1], the first person pronoun "I" refers to Spartacus when Spartacus uses it, and to Antoninus when he uses it. Here's a possible reaction to the example:

> The example shows that distinct tokens of "I" differ in meaning. So the *type* word "I" lacks meaning. It only has a meaning *when used in context*.

This would be an overreaction to the example. There is an important commonality among the semantic interpretations of the different tokens of the first-person singular pronoun: the word "I" always refers to the person using it. Anyone who fails to recognize this feature of the word has not understood it, and so is not a competent user of the word. This suggests that semantic theory should associate a *function* with the word "I," a function which takes a context as its argument and yields as its value the person

speaking in that context. If you like, this function ascribes a meaning to "I" and the referent of each token of the word is the value of the function when evaluated at the relevant context.

Let's look at our second example:

(3) This is a zebra. *[Said by someone while pointing at a zebra]*
(4) This is a zebra. *[Said by someone while pointing at a cleverly decorated mule]*

We can associate with the demonstrative type word "this" a function, which takes a context as argument and yields the object indicated, or otherwise rendered most salient, in the context. Each token of "this" then refers to the value of the function evaluated at the relevant context.

Now for some terminology: the *content* of a token expression is its semantic interpretation *in its particular context*; the *character* of a type expression is a function which maps each context to the appropriate content. Notice that even non-context-sensitive expressions can be said to have characters—in such cases the character is a constant function. Notice also that whole sentences have characters. The character of the sentence "I was hungry yesterday," for example, is a function which maps each context *c* to the proposition that the speaker at *c* was hungry on the day before the time of *c*. The terms "content" and "character" come from David Kaplan, who introduced this way of dealing with context-sensitivity (Kaplan 1989a, 1989b).

This approach to accommodating context-sensitivity is standard for dealing with all of the uncontroversial cases of context-sensitivity, as discussed in section 1. Other (putative) sorts of context sensitivity must be treated differently, as we shall see.

7 Unarticulated Constituents

We now return to an example discussed in section 2.2. Consider:

(17) It's raining. [*Said by John Perry's son, in Palo Alto*]

Plausibly, the sentence "It's raining" is context-sensitive. In some contexts—as in (17)—it expresses the proposition that it is raining in Palo Alto; in other contexts it expresses the proposition that it is raining in Paris. Not everyone will agree that "It's raining" is context-sensitive in this way, but let's put that question to the side, and instead ask how to explain this sort of context-sensitivity, on the assumption that it exists.

Picking up a line of thought from section 6, we might suggest that the source of the context-sensitivity in "It's raining" is that one of the words in the sentence has a non-constant character. "Rain" is the only good candidate, and it seems that the theory will have to go something along these lines:

> The word "rain," when used in a context *c*, refers to the property of raining at the location *contextually salient at c*.

When one uses "It's raining" in a context in which the contextually salient location is Palo Alto, it will express the proposition that it is raining in Palo Alto. On the other hand, when one uses the same sentence in a context in which the salient location

is Paris, it will express a different proposition. So far, so good—the theory seems to work well.

One problem for this approach arises when we consider statements like:

(18) It is raining in Palo Alto.

Suppose we use (18) in a context *c* in which the salient location is Paris. Then, according to the proposed theory, the *content* of "rain" in *c* will be the property of raining in Paris. But then, on the proposed theory, the token of (18) would appear to express the proposition that it is raining in Paris, and it is unclear how "in Palo Alto" could play a role in determining the proposition expressed. Something has clearly gone wrong!

So it seems that we need an alternative to the standard content/character approach. When discussing (17), John Perry famously said:

> In order to assign a truth-value to my son's statement [of (17)] . . . I needed a place. But no component of his statement stood for a place . . . Palo Alto is a constituent of the content of my son's remark, which no component of his statement designated; it is an unarticulated constituent.
>
> (Perry 1986: 206)

Here is Perry's proposal: the underlying relation here is Rain(t, l)—which holds when it is raining at time t in location l. (17) specifies one of the arguments for this relation (the "t" argument), but no syntactic constituent of (17) refers to a location, so the second argument is left, as it were, blank. The sentence, therefore, expresses an incomplete proposition, a proposition with a hole in it (a "propositional radical," as Kent Bach puts it (Bach 1994):

(19) Rain (1st September 1996,_)

Different tokens of (17) express different propositions because this propositional radical gets "filled in" in different ways in different contexts. On this view, (17) expresses the propositional radical (19), but a particular *token* of (17) could express the proposition:

(20) Rain(September 1, 1996, Palo Alto)

Some part of what is asserted in such a case does not correspond to a syntactic constituent of the sentence used to make the assertion. It is, as one says, "unarticulated" (cf. Perry 1986). It is somehow, in a way never made clear in the literature, "introduced" into the proposition expressed without any direction from the uttered sentence.

8 Binding Arguments and Hidden Indexicals

A number of critics have challenged the idea of an unarticulated constituent; in particular, they invoke the so-called binding argument to support the idea that every object in the proposition expressed is put there by an expression in the sentence uttered. We begin with an example. Consider:

(21) Wherever John visits, it's raining.

On its most natural reading, (21) expresses the proposition that for any location x that John visits, it's raining at x. On this reading, some sort of quantification is going on. A natural thought here is that where there is quantification, there must be variable binding. Accordingly, the adverbial phrase "Whenever John visits" is a quantificational phrase, binding a variable in the part of (21) that follows the comma. On this view, the variable does not correspond to any pronounced component of the sentence, but just the same it must be there. It might seem strange to say that there are syntactic components of sentences which are not pronounced, but in fact this is routine in modern syntax.

But what has this got to do with context-sensitivity? Hold onto the idea that "rain" is followed by an unpronounced variable, and look again at (17):

(17) It is raining.

According to the view under discussion, consideration of (21) shows that (17) contains an unpronounced variable. But in the case of (17) there is no quantificational expression that could bind this variable, and so it occurs free. Since it occurs free, its interpretation is fixed contextually. Roughly, in any context of use, the variable will refer to whichever place is most salient in that context of use. On this view, then, the variable in (17) behaves like an indexical (i.e. like "this" or "that"). Because it is not pronounced, it is called a "hidden indexical." Theories of this kind are called "hidden indexical theories."

It has been argued, perhaps most notably by (Stanley and Szabo 2000), that unarticulated constituents theories of (17) (as discussed in the last section) cannot be extended to cope with sentences like (21). If this is right, it provides an argument for hidden indexical theories over unarticulated constituent theories.

It is perhaps worth pausing to draw attention to the key difference between the unarticulated constituents' approach and the hidden indexical approach. Both theories agree that (17), uttered in a certain context, will express the proposition that it is raining in Palo Alto, and both theories agree that no *pronounced* component of (17) refers to Palo Alto. The big difference is that the hidden indexical theory says that there is a component of (17) that refers to Palo Alto—an indexical expression which is present in the syntax but not pronounced—while the unarticulated constituents theory *denies* that there is a component of (17)—even a hidden component—which refers to Palo Alto in the context.

9 The Dynamic Lexicon

It is not uncommon for language users to coin new words "on the fly," or to invent new meaning for old words. (Clark 1983) gives many examples, including these:

> Subjected to the musical equivalent of 72 hours in a dentist's waiting room, Bradley was apparently in real danger of being the first tourist ever Muzakked to death. (From the *San Francisco Examiner*)

> I think that it's across from a quarry. That's the only way I can landmark it. (Said by a person talking about finding a beach).

> I stopped in Perry's for a quick crab. (Herb Caen, meaning a crab that could be eaten quickly)

Examples like these suggest that we should not think that language users just select their words from a fixed lexicon; rather, the lexicon can be modified and extended in the course of conversation. As Ludlow (2000) puts it, the lexicon is dynamic.

Clark uses this point to explain some putative instances of context-sensitivity. One of his examples is that of denominal adjectives—i.e. adjectives derived from nouns, like "Hegelian" (from "Hegel") and "metallic" (from "metal"). Many denominal adjectives have well-established meanings, but others do not. Clark gives this example:

> *Churchillian*, for example, might mean *with a face like Churchill*, *smoking a cigar like Churchill*, *with a speaking style like Churchill*, or any number of other things. In principle, the list is unlimited; in practice, it is limited by what the speaker can assume the addressees know about Churchill, and will be able to see that he is alluding to.

If Clark is right about this, then sentences containing the word "Churchillian" are context-sensitive, but we do not need to account for this by saying that the word is indexical, or by postulating a hidden variable, or by saying that such sentences do not express complete propositions; rather, we should say that the lexicon is dynamic, and the lexical entry for "Churchillian" is different on different occasions.

10 Three Further Supposed Cases of Context-Sensitivity

A summary paper on the literature on context-sensitivity would not be complete without some discussion of contextualism in epistemology, or mention of contextualist solutions to the liar and sorites paradoxes. However, our treatment of these issues will have to be cursory, because of limitations of space.

10.1 Contextualism in Epistemology

Well-known arguments from epistemology—How do you know you're not being tricked by an evil demon?—might lead an epistemologist to say:

(22) I don't know that I am forty years old.

However, when filling in his census form later that day, the very same epistemologist might say:

(23) I know that I am forty years old.

At this point, he is likely to be accused of inconsistency—an accusation that he can attempt to duck by arguing that "know" is context-sensitive. He might say, for example, that there is a shift in which set of epistemic standards are salient between the two contexts.

Contextualism, in epistemology, is the claim that "know" is context-sensitive. Contextualists usually employ this doctrine to reconcile skeptical claims (made in the context of an epistemological discussion) with anti-skeptical claims (made in an everyday context).

10.2 Contextualist Solutions to the Sorites Paradox

Shortly before leaving for the airport with my two suitcases, my traveling companion asks, "Are your suitcases heavy?" In fact, the bags differ in weight by only a gram. I reply:

(24) One of my suitcases is heavy; the other isn't.

It seems that this could not be true: my suitcases hardly differ in weight, so how can I truly say that one is heavy and the other not? The example might lead one to accept:

First: If two things differ only slightly in weight, then either neither is heavy, or they both are.

However, it is notorious that claims like *First* lead to problems. Suppose one has a sequence of suitcases. The first weighs 1kg, the last weighs 100kg, and adjacent suitcases in the sequence differ in weight by only 1g. By repeated application of *First*, one can derive a contradiction from the innocuous assumption that the first suitcase is not heavy and the last is, and the assumption that adjacent bags in the sequence differ only slightly in weight.

Contextualists about vagueness, recalling that there are independent grounds for thinking that "heavy" is context-sensitive (see section 2.3), offer a principle like this in place of *First:*

Second: In a given context, given any two contextually salient objects that differ only slightly in weight, either both objects are in the extension of 'heavy' in the context or neither is.

To put it roughly, the line that divides heavy things from the things which are not heavy moves around so that it never lies between two objects that are contextually salient and very similar in weight. The reason why it is improper for me to assert (24) is that both of the suitcases are contextually salient, and they differ only slightly in weight.

The theory is attractive because it rejects *First* (which leads to contradiction) and also offers an explanation of why *First* is so attractive. Vague predicates appear to be "boundaryless" because the boundary moves in such a way that it never lies between two contextually salient objects. Wherever you look, there it isn't.

10.3 Contextualist Solutions to the Liar

Consider sentence (25):

(25) (25) is not true.

What is its semantic status? We have, in answering this question, at least the following three options:

Option One: (25) is true.
Option Two: (25) is not true.
Option Three: (25) fails to express a proposition; it lacks a truth-condition.

Prima facie, Options One and Two are inconsistent, as are Options One and Three. It turns out that it is difficult to adopt one of these options *and then stick with it despite further thought*. Whichever option one chooses, one will find oneself inclined to move to a different one.

One might start by reasoning as follows. Suppose for contradiction that (25) expresses a proposition. Then:

"(25) is not true" is true if and only if (25) is not true.

But "(25) is not true" just is (25), so by Leibniz's law we can infer:

(25) is true if and only if (25) is not true.

This is a contradiction—so by *reductio* we can infer that (25) does not express a proposition. This argument gives us a good reason for accepting Option Three.

However, further thought suggests otherwise. We seem to have shown that (25) does not express a proposition. But it would seem to follow that (25) is not true (since only a sentence that expresses a proposition can be true). So we should accept Option Two.

Again, further thought will force us to change our mind. We seem to have given a sound argument for accepting Option Two—i.e. for accepting:

(25) is not true.

But this just is the sentence (25), so we have given a sound argument for (25). It would seem to follow that (25) is true after all, and we then conclude that we should accept Option One. Repeating our initial line of argument will take us back to Option Three—and on it goes.

So to repeat, whatever semantic status we ascribe to (25), further thought seems to force us to assign it a different status. The contextualist concludes that (25) must be *changing* status; on this view, (25) is context-sensitive; and as we reason, the context changes in such a way that (25) changes status.

11 Summary

- Some words—including "I," "this," "current" and "today"—are uncontroversially context-sensitive.
- It is controversial how far context-sensitivity extends beyond these 'easy' examples.
- At one extreme, minimalists argue that the only context-sensitivity in natural language is that introduced by the 'easy' examples. At the other extreme, radical contextualists argue that examples of context-sensitivity are so widespread and multifarious that they cannot be accounted for by a systematic semantic theory.
- Radical contextualist positions are hard to state coherently (see section 4), and are at any rate at odds with the results of tests for context-sensitivity (see section 5).
- We have discussed several different ways of explaining context-sensitivity: Kaplan's character/content approach (section 6); the 'unarticulated constituents' approach (section 7); the 'hidden indexicals' approach (section 8) and the 'dynamic lexicon'

approach (section 9). The character/content approach is standard for the 'easy' examples of context-sensitivity discussed in section 1, while the other approaches are used in more controversial cases.

- It has been argued that postulating context-sensitivity can help us with the liar paradox and the sorites paradox, and that it can be a useful part of a response to skepticism in epistemology.

References

Bach (1994) "Conversational Impliciture," *Mind and Language* 9: 124–162.

Beall J. and M. Glanzberg (Forthcoming) 'The Liar Paradox' in *The Stanford Encyclopedia of Philosophy*.

Bezuidenhout, A. (2002) "Radical Pragmatics" in J. K. Campbell et al. (eds.) *Meaning and Truth: Investigations in Philosophical Semantics*. New York: Seven Bridges Press, pp. 292–302.

Borg, E. (2004) *Minimal Semantics*. Oxford: Oxford University Press.

Burge, T. (1979) "Semantical Paradox," *Journal of Philosophy* 76: 169–198.

Cappelen, H. and J. Hawthorne (2007) "Locations and Binding," *Analysis* 67: 95–105.

—— (2009) *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Cappelen, H. and E. Lepore (2002) "Indexicality, Binding, Anaphora and A Priori Truth," *Analysis* 62: 271–281.

—— (2005) *Insensitive Semantics*. Oxford: Basil Blackwell.

Carston, R. (2002) *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Clark, H. (1983) 'Making Sense of Nonce Sense' in Clark (1992) and Flores d'Arcais and Jarvella (1983)

—— (1992) *Arenas of Language Use*. Chicago: University of Chicago Press.

DeRose, K. (1995) "Solving the Sceptical Problem," *Philosophical Review* 104: 1–52.

Devault, D. and M.D. Stone (2006) "Scorekeeping in an Uncertain Language Game," The 10th Workshop in the Semantics and Pragmatics Dialogue (brandial 2006), University of Potsdam, Germany.

Fara, D.G. (2000) "Shifting Sands: An Interest-Relative Theory of Vagueness", *Philosophical Topics*, 28: 45–81. Published under the name 'Delia Graff'.

Flores D'Arcais, G.B. and R.J. Jarvella (1983) *The Process of Understanding Language*. New York: Wiley.

Grice, H.P. (1975) "Logic and Conversation" in P. Cole and J. Morgan (eds.) *Syntax and Semantics, 3: Speech Acts*. New York: Academic Press. Reprinted in H. P. Grice (ed.) *Studies in the Way of Words*, pp. 22–40. Cambridge, MA: Harvard University Press (1989).

Haegeman, L. (1994) *Introduction to Government and Binding Theory*, 2nd edition. Cambridge, MA: Wiley-Blackwell.

Hamm, F., H. Kamp and M. Lambalgen (2006) "There is No Opposition between Formal and Cognitive Semantics," *Theoretical Linguistics* 32: 1–40.

Jaszczolt. K.M. (ed.) (2000) *The Pragmatics of Propositional Attitude Reports*. Oxford/New York: Elsevier Science.

Kaplan, David (1989a) "Demonstratives" in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, pp. 481–563.

—— (1989b) 'Afterthoughts' in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, pp. 565–614.

Keefe, R. and P. Smith (eds.) (1996) *Vagueness: A Reader*. Cambridge, MA: MIT Press.

Lewis, D. (1979) "Scorekeeping in a Language Game," *Journal of Philosophical Logic*, 8: 339–359.

—— (1996) "Elusive Knowledge," *Australasian Journal of Philosophy* 74: 549–567.

Ludlow, P. (1999) *Semantics, Tense, and Time: An Essay in the Metaphysics of Natural Language*. Cambridge, MA: MIT Press.

—— (2000) "Interpreted Logical Forms, Belief Attribution, and the Dynamic Lexicon," in Jaszczolt (2000).

Parsons, C. (1974) "The Liar Paradox," *Journal of Philosophical Logic* 3: 381–412.

Perry, J. (1986) "Thought without Representation," *Proceedings of the Aristotelian Society Supplementary* 60: 263–283. Reprinted in Perry (1993).

—— (1993) *Problem of the Essential Indexical and Other Essays*. Oxford: Oxford University Press.

—— (1998) "Indexicals, Contexts and Unarticulated Constituents" in *Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference*. Stanford: CSLI Publications.

Raffman, D. (1996) "Vagueness and Context-Sensitivity", *Philosophical Studies* 81: 175–192.

Recanati, F. (2004) *Literal Meaning*. Cambridge: Cambridge University Press.
Roberts, C. (2004) "Context in Dynamic Interpretation" in L. Horn and G. Ward. (eds.) *The Handbook of Pragmatics*. Oxford: Blackwell, pp. 197–220.
Sainsbury, M. (1996) "Concepts without Boundaries" in Keefe and Smith (1996), pp. 251–264.
Searle, John, (1978) "Literal Meaning," *Erkenntnis*, 13: 207–224.
—— (1980) "The Background of Meaning" in J. Searle, J. F. Kiefer and M. Bierwisch (eds.) *Speech Act Theory and Pragmatics*. Dordrecht: Reidel, pp. 221–232.
—— (1983) *Intentionality*. Cambridge: Cambridge University Press.
Shapiro, S. (2006) *Vagueness in Context*. Oxford: Oxford University Press.
Soames, S. (1999) *Understanding Truth*. Oxford: Oxford University Press.
Stalnaker, R. (1978) "Assertion" in P. Cole (ed.) *Syntax and Semantics Vol. 9*. New York: New York Academic Press, pp. 315–332.
—— (1998) "On the Representation of Context," *Journal of Logic, Language and Information*, Special Issue on Context in Linguistics and Artificial Intelligence, January, 7 (1): 3–19.
Stanley, J. (2002) "Making it Articulated," *Mind and Language* 17 (1&2): 149–168. Reprinted in Stanley (2007).
—— (2005) *Knowledge and Practical Interests*. Oxford: Oxford University Press.
—— (2007) *Language in Context: Selected Essays*. Oxford: Oxford University Press.
Stanley, J. and Z. Szabo (2000) "On Quantifier Domain Restriction," *Mind and Language* 15 (2): 219–261.
Steedman, M. and M. D. Stone (2006) "Is Semantics Computational?" *Theoretical Linguistics* 32 (1): 73–89
Travis, Charles (1985) "On What is Strictly Speaking True," *Canadian Journal of Philosophy* 15 (2): 187–229. Reprinted in Travis (2008)
—— (1989) *The Uses of Sense*. Oxford: Oxford University Press.
—— (1996) "Meaning's Role in Truth,"*Mind* 100: 451–466.
—— (1997) "Pragmatics" in B. Hale and C.Wright (eds.) *A Companion to the Philosophy of Language*. Oxford: Oxford University Press.
—— (2008) *Occasion Sensitivity: Selected Essays*. Oxford: Oxford University Press.
Westerstahl, Dag (1985) "Determiners and Contexts Sets" in J. van Benthem and A. ter Meulen (eds.), *Quantifiers in Natural Languages*. Dorgrecht: Foris, pp. 45–71.
—— (1989) "Quantifiers in Formal and Natural Languages" in D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic*, IV (Dordrecht: Kluwer), pp. 1–131.

1.11

RELATIVISM

John MacFarlane

Relativism, in the sense at issue here, is the view that some assertions or beliefs can be characterized as accurate, or objectively correct, only relative to a context of assessment. We will begin by looking at a few cases that make relativist views tempting, then consider the technical and philosophical issues such views raise.

Motivation

Epistemic Modals

"Joe might be in Boston," says Maria. In doing so, she expresses a certain state of mind, making it clear that she does not take her information to rule out Joe's being in Boston. But does she *assert* this? There are reasons to think not. George, who knows that Joe is just down the hall, might disagree with what Maria has said. And, after Maria learns that Joe is in fact in Berkeley, she too may think that she was wrong, and retract her claim: "I said he might be in Boston, but that's wrong—he can't be in Boston, since George just saw him down the hall."

Noticing these things, one might conclude that Maria hasn't asserted anything but has only expressed her uncertainty. But suppose one does want to think of her speech act as an assertion. What are the options? One could explain the disagreement and retraction facts by taking her to have made a claim about what is left open by the information of a broader group, including her possible interlocutors. But then it becomes hard to understand how she could have taken herself to be warranted in making this claim. We need to understand how it is that she can legitimately assert that Joe might be in Boston on the basis of nothing more than her own ignorance, while others—and her own later self—can evaluate what she has said in relation to *their* information.

All of this could be explained if we could make sense of the idea that the accuracy of Maria's assertion is not an absolute matter but varies with perspective. We could then say that it is the *assessor's* information, not the speaker's, that matters for the accuracy of an epistemic modal claim. That would explain why, when Maria learns that Joe is right down the hall, she takes her original claim that he might be in Boston to have been inaccurate, and retracts it. In this way we might capture the subjectivity of epistemic modal claims—the way in which their accuracy depends on what is known, not on any objective feature of the world—without construing them (as contextualist views do) as claims about what some particular person or group knows. (See Egan et al. 2005, Egan 2007, Stephenson 2007, MacFarlane 2011; for critical remarks, see Wright 2007, von Fintel and Gillies 2008, Dietz 2008.)

Taste Predicates

"This vegemite sandwich is tasty," says Joe. In doing so, he expresses his liking for the sandwich, but does he do anything more? Does he assert something, and if so, on what does the accuracy of his assertion depend?

Contextualists hold that Joe has asserted something about the relation between the food and himself, or between the food and a certain body of tastes. But then it is puzzling why Sasha, whose tastes are very different, should reject Joe's assertion. On a relativist view, by contrast, Joe's assertion can only be classed as accurate or inaccurate relative to a context of assessment, and it is the assessor's tastes that matter for its accuracy, rather than the speaker's. If Sasha despises the taste of vegemite, while Joe loves it, then Joe's assertion will be inaccurate as assessed by Sasha but accurate as assessed by Joe. And, if Joe later comes to hate the taste of vegemite, his earlier belief will be inaccurate relative to his current context, and he will rightly consider it to have been wrong. In this way, the relativist tries to capture the subjectivity of taste claims—the way in which their accuracy depends on a sense of taste—without reducing them to claims about how various foods relate to the tastes of a person or group. (See Kölbel 2002, Lasersohn 2005, 2009, Stephenson 2007, MacFarlane 2007; and for criticism, Glanzberg 2007, Cappelen and Hawthorne 2009, ch. 4.)

Future Contingents

"There will be a sea battle tomorrow," says Themistocles, at a time when it is objectively indeterminate whether there will be a sea battle the next day. Is his assertion accurate or inaccurate? The question can only be answered, according to a relativist view, relative to a particular context of assessment. Themistocles' assertion is inaccurate as assessed from the moment at which it is made (m_0), accurate as assessed from a moment one day later on a possible future history with a sea battle (m_1), and inaccurate as assessed from a moment one day later on a possible future history without a sea battle (m_2). In this way the relativist splits the difference between the supervaluationist, who holds that the assertion is (absolutely) inaccurate—and has trouble explaining why it need not be retracted by the asserter at m_1—and the "true futurist," who holds that the assertion is either accurate or inaccurate absolutely, even though no objective feature of the situation breaks the symmetry between the two possible histories. (See MacFarlane 2003, 2008; and for criticism, Heck 2006.)

Such proposals raise two foundational questions in semantics:

> *The technical question* What revisions are needed to standard truth-conditional semantic frameworks to make room for relativist accounts of this kind?
>
> *The philosophical question* How can we make philosophical sense of the idea that the accuracy of an assertion or belief is assessment-relative? Do we really understand what such proposals *say*? How do relativist theories differ from non-relativist theories in their predictions about language use?

In what follows, we will briefly consider how each of these questions might be answered.

The Technical Question

Index and Context

The task of a semantic theory, according to Lewis (1980), is to define truth at a context of use for sentences of a language *L*. Why is this our target notion? Because if we know the conditions under which sentences of *L* are true at a context, we know enough to understand and communicate with speakers of *L*. One counts as a speaker of *L* in virtue of participating in a convention of uttering only true sentences of *L* and trusting other *L*-speakers to do the same. Thus, if we know that a speaker of *L* utters *S* in context *c*, and we know that in *L*, *S* is true at *c* if and only if *p*, we can presume that the speaker believes that *p* and intends for us to believe this too. And if we want to communicate that *q* to another speaker, we have only to select a sentence *T* that is mutually known to be true in our context if and only if *q*.

Because languages contain infinitely many sentences, and we want a finite characterization of our semantic knowledge, our definition of 'true at *c*' must be recursive. However, it is not in general possible to state the condition for a sentence to be true at a context in terms of the conditions for its constituents to be true at a context. One reason is that the constituents of some complex sentences are not sentences but open formulas. Another reason is that a proper treatment of modal and temporal operators requires considering the truth value of the embedded sentence at times and worlds other than the time and world of the context. For example, to evaluate ⌜Yesterday: ϕ⌝ we need to consider the truth value of ϕ relative to the day before the present day. If we are recursively defining truth at a context, we will need to evaluate ϕ at a context *c′* that differs from our current context c only in the time of the context. However, as Lewis points out (1983, 29), there is no guarantee that there will be such a context. Contexts represent possible occasions of use, and are not arbitrary collections of parameters. Since the agent of the context must exist at the time and world of the context, if we shift the time back to before the agent of the context existed, there will be no corresponding context. Moreover, even if there were a context *c′* that differed from *c* only in the time of the context, we would not want to say that ⌜Yesterday: ϕ⌝ is true at *c* iff ϕ is true at *c′*. For ϕ might contain temporal indexicals like 'today' or 'now', and these must be evaluated with respect to the original context *c*. 'Yesterday it was warmer than it is now' can be true, even though the embedded sentence 'It is warmer than it is now' is not true relative to *an*y context (Kamp 1971, Kaplan 1989).

The standard solution to these problems is to recursively define truth at a context and an *index*, which is simply a collection of independently shiftable parameters (world, time, assignment of values to the variables, and so on). The clause for 'Yesterday' could be:

> *Yesterday*⌜Yesterday: ϕ⌝ is true at *c*, (w, t, a) iff ϕ is true at *c*, $\langle w, t', a\rangle$, for some t' belonging to the day before *t*.

Although what we really care about at the end of the day is truth at a context, a recursive definition of truth at a context and index can serve our purposes, as long as we can recover a definition of truth at a context from it. And we can:

> *Lewisian postsemantics* A sentence ϕ is true at c iff ϕ is true at *c*, $\langle w_c, t_c, a\rangle$, for any assignment *a*. (Here w_c and t_c are the world and time of the context *c*.)
>
> (cf. Lewis 1983, 31)

MacFarlane (2003, 2005a) calls this step the *postsemantics*, to distinguish it from the recursive definition of truth at a context and index, which we might call the *compositional semantics*.

Assessment Sensitivity

As Lewis makes clear, the role of the index in such a theory is entirely technical. If there is no operator that shifts a parameter, then the parameter is not needed in the index. And, if there *is* an operator that shifts the parameter, no further justification is needed for putting that parameter in the index. The index is just part of the machinery for giving a systematic definition of truth at a context; it plays no further role and is subject to no further extrasystematic constraints.

So the presence of outlandish parameters of the index (taste, information state, etc.) does not itself make a semantic theory "relativist" in any philosophically interesting sense. It does not prevent the theory from assigning truth values to sentences at contexts, or making absolute judgements of the accuracy of assertions.

We can see this more clearly if we change our target notion to make room for the possibility that accuracy is not absolute. Instead of defining truth at a context of use, let us define truth at a context of use and *context of assessment*—a possible situation in which a speech act or mental attitude might be assessed. We can then distinguish two kinds of contextual sensitivity. An expression is *use-sensitive* if its extension depends on features of the context of use, and *assessment-sensitive* if its extension depends on features of the context of assessment. Or, making finer discriminations: an expression is *F-use-sensitive* if its extension depends on the *F* of the context of use, and *F-assessment-sensitive* if its extension depends on the *F* of the context of assessment. A "relativist" semantic theory is one that takes some expressions to be assessment-sensitive. This amounts to taking accuracy to be relative to contexts of asssement, since

> *Accuracy* An assertion of *S* at *c* is *accurate*, as assessed from *c′*, iff *S* is true as used at *c* and assessed from *c′*.

A compositional semantics that employs a taste parameter in the indices (perhaps to make sense of an operator like 'on any standard of taste') need not take any expression to be taste-assessment-sensitive. For such a semantics can be combined with either a contextualist or a relativist postsemantics:

> *Contextualist postsemantics* *S* is true as used at *c* (and assessed from *c′*) iff *S* is true at *c*, $\langle w_c, t_c \rangle$, where w_c is the world of *c* and t_c is the taste of the agent at *c*.
>
> *Relativist postsemantics* *S* is true as used at *c* and assessed from *c′* iff *S* is true at *c*, $\langle w_c, t_{c'} \rangle$, where w_c is the world of *c* and $t_{c'}$ is the taste of the agent at *c′* (i.e., the assessor).

According to the contextualist theory, the accuracy of an assertion of 'This is tasty' will depend on the physical disposition of the demonstrated food at the time of the assertion and on the speaker's tastes at the time of the assertion. According to the relativist theory, by contrast, it will depend on the physical disposition of the demonstrated food at the time of the assertion and the *assessor's* tastes at the time of assessment. Since the same assertion can be assessed from indefinitely many perspectives, there are only perspective-relative answers to the question whether the assertion was accurate.

Content and Circumstance

Lewis (1980) argued that, for purposes of doing semantics, one need not talk about contents or propositions at all. In that respect he differs from Kaplan (1989), who adopts a two-stage picture of semantics. In the first stage of such a semantics, we define the *expression* relation, specifying for any sentence *S* and context *c* the *structured proposition* that *S* expresses at *c*. In the second stage, we give a recursive truth definition of truth at a *circumstance of evaluation* for the structured propositions. In Kaplan's framework, circumstances are time/world pairs, but Kaplan acknowledges that "other features," such as location, may also be needed (1989, 504). Though it is not clear that Kaplan would go so far, one could in principle take subjective features like *taste* or *body of information* to be parameters of circumstances.

The truth of a sentence relative to a context is the joint product of these two stages:

> *Kaplanian postsemantics* A sentence *S* is true at a context *c* iff the content expressed by *S* at *c* is true at the circumstance of *c*, that is, $\langle w_c, t_c \rangle$.
>
> (cf. Kaplan 1989, 522)

Here, the context plays two distinct roles in determining a truth value for an occurrence of a sentence. It plays a *content-determinative* role in fixing which content is expressed, and a *circumstance-determinative* role in fixing at which circumstance this content is to be evaluated. Hence, there will in general be two distinct ways to take some expression to be *F*-use-sensitive. One can take the expression's *content* to depend on the *F* of the context of use, or one can take the *F* of the context of use to play a role in picking out the "circumstance of the context"—the circumstance one looks at in determining whether the sentence is true at the context, and in determining whether an assertion or belief in the context is accurate.

Contextualist views typically use the first model. Contextualists about 'tasty', for example, often hold that the sentence 'vegemite is tasty' expresses different propositions depending on the tastes of the speaker. But the second model is also an option. One could take the content of 'vegemite is tasty' to have truth values relative to worlds, times, and *tastes*, and let the tastes of the speaker (or more generally the taste relevant at the context of use) help determine which triple is the "circumstance of the context." On such a view, which following MacFarlane (2009) we can call *nonindexical contextualism*, sentences have truth values relative to contexts of use, and accuracy is absolute, not assessment-relative. An assertion that vegemite is tasty is accurate, simpliciter, just in case the taste of vegemite is good according to the taste of the speaker (or whatever taste is relevant at the context of use). On this view, 'vegemite is tasty' is taste-use-sensitive, not taste-assessment-sensitive.

Some writers in the literature use the term "truth relativism" for any view that relativizes propositional truth to "nonstandard" features like tastes or information states. However, as the possibility of nonindexical contextualist views shows, such relativism at the level of circumstances of evaluation is compatible with the traditional view that the accuracy of particular assertions and beliefs is an absolute matter. (Conversely, one can have assessment sensitivity without any nonstandard parameters of circumstances. For an example in which circumstances are just worlds, see MacFarlane 2008.) Since most of the traditional worries about truth relativism are worries about the rejection of this

traditional view, and not about the individuation of propositions, we follow MacFarlane (2005a) in reserving the label "truth relativism" for views that countenance assessment sensitivity. (Lasersohn 2005 does not explicitly relativize truth to contexts of assessment, so it might appear that his view is not relativist, in the sense distinguished here. However, that would be misleading. For Lasersohn, a "context of use" is not a concrete possible situation in which a sentence might be used but an abstract sequence of parameters. How we should set these parameters in interpreting an utterance of 'This is tasty' depends, Lasersohn holds, on both features of the concrete speech situation, which help determine the reference of 'this' and the relevant world and time, and on features of the concrete assessment situation, which determine the "judge" of the context. So both the concrete use situation and the concrete assessment situation play a role, even though the distinction between them is not made notationally salient.)

If we want a view on which 'vegemite is tasty' is taste-assessment-sensitive, then again there are two options in a Kaplanian framework. If we give the context of assessment a content-determinative role, then we have a form of *content relativism*.

> *Content-relativist postsemantics* S is true as used at c and assessed from c' iff the content expressed by S at c, as assessed from c', is true at the circumstance of c, that is, $\langle w_c, t_c \rangle$, where w_c and t_c are the world and taste of c.

On a content-relativist view, there is no absolute answer to the question which proposition is expressed by a sentence at a context of use. Different propositions are expressed, relative to different contexts of assessment. (Content relativism has had few advocates in the literature, as compared with truth relativism. But see Cappelen 2008 and Weatherson 2009.)

If we give the context of assessment a circumstance-determinative role, then we get a

> *Truth-relativist postsemantics* S is true as used at c and assessed from c' iff the content expressed by S at c is true at the circumstance $\langle w_c, t_{c'} \rangle$, where w_c is the world of c and $t_{c'}$ is the taste of c'.

Here the context of use and context of assessment jointly determine which circumstance is relevant to the truth of the sentence relative to these contexts. According to this view, an assertion or belief that something is tasty will be accurate, as assessed from another context c', if the way the thing tastes at the time of the assertion or belief is good according to the tastes of the assessor at c'. So the very same assertion or belief may be accurate, as assessed by one observer, and inaccurate, as assessed by another. This is one way of trying to capture the idea that tastiness depends on our evaluative reactions, without reducing all ascriptions of tastiness to claims about how foods strike some particular person or group.

Variables

A relativist who adopts a Lewisian or Kaplanian approach can keep the idea that an adjective like 'tasty' is semantically a one-place predicate. There is no extra argument place for a taste or judge. This may be counted an advantage, but it makes accounting for binding phenomena more awkward. We must analyze

(1) That's tasty to Sam

as

(2) To Sam: tasty(that),

where 'To Sam:' is an operator that shifts the taste parameter of the index (or, in Lasersohn's 2005 version, a predicate-modifier). And we must analyze

(3) Everyone got something tasty

(on the reading where it means that everyone got something tasty to them) as

(4) [Everybody x][Something y](x got y & to x: tasty(y)).

A tempting alternative is to treat 'tasty' as semantically *two-place*, with an implicit argument place for a taste. We could then analyze (1) as

(5) Tasty(that, Sam's taste)

and (3) as

(6) [Everybody x][Something y](x got y & tasty(y, x's taste))

Following Stephenson (2007), we could take the extra argument place to be filled, when not bound, by a special unpronounced pronomial element, PRO_J. On Stephenson's account, PRO_J denotes the judge of the index; here, we could take it to denote the taste of the index, or (if we had no other need for a taste parameter in the index) we could take its value to be set directly by the taste of the context of assessment.

Thus, although it is sometimes assumed that relativist views require operators to motivate novel parameters in the index (Cappelen and Hawthorne 2009, ch. 3, Stanley 2005, ch. 7), the issue of assessment sensitivity can be pried apart from questions about the parameters of indices, and from related questions about whether there are operators that shift these parameters. This should not be surprising. As we will see below, assessment sensitivity is a way to explain certain features of the use of nonembedded occurrences of sentences. Thus it ought to be largely independent of the issues about *embedded* occurrences of expressions that motivate indices.

The Philosophical Question

What motivates work on assessment sensitivity is an

> *Empirical claim* There are assessment-sensitive expressions in natural languages. Some of the things we say and think are assessment-sensitive.

Defending this claim would require detailed examination of the use of particular bits of language and careful consideration of alternative explanations. However, the empirical claim presupposes a weaker, conceptual claim that can be established—if it can be established at all—through philosophical clarification:

Conceptual claim The possibility of assessment-sensitive expressions is coherent and intelligible. We know what it would be for an expression to be assessment-sensitive, whether or not this possibility is realized in natural languages.

The ground-clearing of the last section has helped us see what assessment sensitivity is not, but it has not yet told us what it is. If we are given a definition of *truth at a context of use* for a language, then we have a reasonable grip on how its expressions are to be used, since we know that (to a first approximation) speakers try to assert things that are true at their current contexts, and expect others to do the same (see Dummett 1981, 302, Lewis 1980, §2, 1983, §III). But what if we are given a definition of *truth at a context of use and context of assessment*? How does this notion connect to the use of language? What is the significance of assessment sensitivity for our linguistic practices? Something must be said here, or it is not clear that in talking of truth relative to a context of use and context of assessment, we are saying anything meaningful at all.

It is important to realize that both 'true as used at a context *c*' and 'true as used at *c* and assessed from *c''* are theoretical terms. Ordinary speakers don't talk of sentences being true relative to contexts. When they say things like, 'What Joe said is true' and 'None of the claims in Sally's report are true', they are using a monadic predicate that applies to propositions. So, we need to say something about what these relativized truth predicates mean if we are to use them in theories of meaning. It might seem tempting to define 'is true as used at *c*' as 'would express a truth if it were used at *c*', but this doesn't quite work: It would render the sentence 'An English sentence is being used' true at all contexts, for example. In addition, if the ordinary monadic truth predicate is nothing more than a logical device for expressing generalizations, as many philosophers have claimed, then a definition that fixes its extension cannot serve as a theory of meaning (see Dummett 1959 for the point, Williams 1999 for dissent, and Patterson 2005 for a recent defense). A more promising approach is to let the meanings of these technical terms be fixed (as with many technical terms) by their role in a larger theory—here, a theory of language use. If we understand what is implied about the use of language by claims made using the relativized truth predicate, then we understand this predicate well enough.

In the case of 'true as used at *c*', we can get pretty far by saying, with Dummett and Lewis, that speakers understand each other (in the absence of special reasons) to be trying to assert only truths; that is, they follow a norm

T1 Assert ϕ in *c* only if ϕ is true as used at *c*.

Given that knowledge, plus a definition of truth at a context of use for a language, you will be in a position to understand utterances others make, and to communicate your own thoughts to others using the language. Simplifying somewhat: When you want to communicate that you are hungry, you will find a sentence whose truth in your current context depends on your being hungry; when someone else utters a sentence that your definition tells you is true in their context just in case they are hungry, you will take them to be communicating that they are hungry.

But suppose we have a doubly relativized truth predicate: 'true as used at *c* and assessed from *c''*'? What shall we say about how ascriptions of assessment-relative truth relate to language use?

We might try quantifying over contexts of assessment in our norm for assertion:

> **T2** Assert ϕ in c only if ϕ is true as used at c and assessed from c', for some/most/all contexts of assessment c'.

Or, more plausibly, we might simply privilege a particular context of assessment—the one the speaker is in when making the assertion:

> **T3** Assert ϕ in c only if ϕ is true as used at c and assessed from c.

But if this is *all* we say, assessment sensitivity will be an idle wheel, with no real use in semantic theorizing. For, given any definition T of 'true as used at c and assessed from c'' for a given language, according to which certain expressions are assessment-sensitive, we can construct another definition T^* that yields the same norms for language use as T does (when combined with **T3**) but does not make any expressions assessment-sensitive. Here's the recipe:

> **T*** ϕ is true as used at c and assessed from c' iff according to T, ϕ is true as used at c and assessed from c.

T3 cannot help us distinguish these two definitions, because T3 only tells us the significance of the assessment-relative truth predicate for the special case where the context of assessment is the same as the context of use, and in such cases T and T^* will always agree. If T3 is all we have to say to connect the assessment-relative truth predicate to language use, then, we will never have evidence that favors T over T^*, so we will never have good reason to embrace assessment sensitivity (see MacFarlane 2005a, which draws on Evans 1985, 348).

However, we can distinguish theories that posit assessment sensitivity from theories that do not if we consider not just norms for making assertions but norms for retracting, challenging, and/or justifying them. For, when one is considering a challenge to an assertion, or considering whether to retract an assertion made earlier, there are potentially two relevant contexts: the context in which the assertion was made and the context in which it is now being assessed. There is room here, then, for the context of assessment to play a significant role in the norms for these speech acts.

For example, suppose we combine T3 with the following norm for retraction:

> **R** Retract at c' an assertion of ϕ made at c if ϕ is not true as used at c and assessed from c'.

Now we can distinguish between relativist and nonindexical contextualist views about 'tasty'. Both will say that one may assert 'this is tasty' at c only if the demonstrated food is good according to one's tastes at c. But, given R, the two views will differ about the condition under which such an assertion must be retracted at some later context c' where one's tastes are different. According to the nonindexical contextualist version, 'this is tasty' is not assessment-sensitive; so it is true as used at c and assessed from c' just in case it is true as used at c and assessed from c. If it was okay to make the assertion in the first place, then, it need not be retracted. According to the relativist view, however, the truth of 'this is tasty' (at a context of use and context of assessment) depends on the taste of the context of assessment; so, even if 'this is tasty' is true as used at c and assessed from c, it may be false as used at c and assessed from c'. In such a case, the assertion would have to be retracted, even though one did not violate any norm in making it.

Consider a concrete example on which the relativist and nonindexical contextualist views differ. Suppose that at age ten, I asserted 'fish sticks are tasty'. Now my tastes have changed; I find fish sticks disgusting, and I deny that fish sticks are tasty. Am I obligated to retract my earlier assertion? The nonindexical contextualist says that I am not. I can let this earlier assertion stand, since, although it was not not specifically *about* my tastes, it concerned my tastes at the time. The relativist says that I must retract; I cannot coherently reject the proposition I asserted as false while holding fast to my earlier assertion.

Of course, claims about what speakers are obligated to do are not themselves claims about what speakers *will* do. Still, this is enough to give us an empirical grip on the distinction. After all, speakers can be presumed to be trying to satisfy the norms, except when there is good reason not to, and they may have intuitions about their obligations to retract. If assessment-sensitive and non-assessment-sensitive theories make different predictions about proprieties for language use, they will also make different predictions about language use. Arguably, this is enough to vindicate the conceptual claim, and to tell us what to look for in evaluating the empirical claim. (See MacFarlane 2003, 2005a, 2007 for the general strategy pursued here.)

References

Cappelen, H. (2008) "Content Relativism and Semantic Blindness," in García-Carpintero and Kölbel (2008), 265–286.

Cappelen, H. and J. Hawthorne (2009) *Relativism and Monadic Truth* (Oxford: Oxford University Press).

Dietz, R. (2008) "Epistemic Modals and Correct Disagreement," in García-Carpintero and Kölbel (2008), 239–263.

Dummett, M. (1959) "Truth," *Proceedings of the Aristotelian Society* n.s. 59: 141–62.

Dummett, M. (1981) *Frege: Philosophy of Language*, 2nd ed. (Cambridge, MA: Harvard University Press).

Egan, A. (2007) "Epistemic Modals, Relativism, and Assertion," *Philosophical Studies*, 133: 1–22.

Egan, A., J. Hawthorne, and B. Weatherson (2005) "Epistemic Modals in Context," in G. Preyer and P. Peter (eds.) *Contextualism in Philosophy* (Oxford: Oxford University Press).

Evans, G. (1985) "Does Tense Logic Rest on a Mistake?" in *Collected Papers* (Oxford: Oxford University Press), 343–63.

García-Carpintero, M. and M. Kölbel (eds.) (2008) *Relative Truth* (Oxford: Oxford University Press).

Glanzberg, M. (2007) "Context, Content, and Relativism," *Philosophical Studies* 136: 1–29.

Heck, R. (2006) "MacFarlane on Relative Truth," *Philosophical Issues* 16: 88–100.

Kamp, H. (1971) "Formal Properties of 'Now'," *Theoria* 37: 227–273.

Kaplan, D. (1989) "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in J. Almog, J. Perry and H. Wettstein (eds.) *Themes from Kaplan* (Oxford: Oxford University Press), 481–566.

Kölbel, M. (2002) *Truth Without Objectivity* (London: Routledge).

Lasersohn, P. (2005) "Context Dependence, Disagreement, and Predicates of Personal Taste," *Linguistics and Philosophy* 28: 643–86.

Lasersohn, P. (2009) "Relative Truth, Speaker Commitment, and Control of Implicit Arguments," *Synthese* 166: 359–74.

Lewis, D. (1980) "Index, Context, and Content," in Stig Kanger and Sven Öhman (eds.) *Philosophy and Grammar* (Dordrecht: Reidel). Reprinted in Lewis (1983).

Lewis, D. (1983) *Philosophical Papers*, vol. 1 (Oxford: Oxford University Press).

MacFarlane, J. (2003) "Future Contingents and Relative Truth," *Philosophical Quarterly* 53: 321–36.

MacFarlane, J. (2005a), "Making Sense of Relative Truth," *Proceedings of the Aristotelian Society* 105: 321–39.

MacFarlane, J. (2007) "Relativism and Disagreement," *Philosophical Studies* 132: 17–31.

MacFarlane, J. (2008), "Truth in the Garden of Forking Paths," in García-Carpintero and Kölbel (eds.) 2008, 81–102.

MacFarlane, J. (2009) "Nonindexical Contextualism," *Synthese* 166: 231–50.

MacFarlane, J. (2011) "Epistemic Modals are Assessment Sensitive." in E. Egan and B. Weatherson (eds.) *Epistemic Modality* (Oxford: Oxford University Press).

Patterson, D. (2005) "Deflationism and the Truth Conditional Theory of Meaning," *Philosophical Studies* 124: 271–94.
Stanley, J. (2005) *Knowledge and Practical Interests* (Oxford: Oxford University Press).
Stephenson, Tamina (2007) "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste," *Linguistics and Philosophy* 30: 487–525.
von Fintel, K. and A. Gillies (2008) "CIA Leaks," *Philosophical Review* 117: 77–98.
Weatherson, B. (2009) "Conditionals and Indexical Relativism," *Synthese* 166: 333–57.
Williams, M. (1999) "Meaning and Deflationary Truth," *Journal of Philosophy* 96: 545–64.
Wright, C. (2007) "New Age Relativism and Epistemic Possibility: the Question of Evidence," *Philosophical Issues* 17: 262–83.

Further Reading

MacFarlane, J. (2005b) "The Asssessment Sensitivity of Knowledge Attributions," in T. Szabó Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology* 1: 197–233. (Argues that a relativist treatment of knowledge attributions combines the virtues of contextualism and invariantism.)
Recanati, F. (2007) *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford University Press. (An extended argument for what is here called non-indexical contextualism, with applications.)
Richard, M. (2004) "Contextualism and Relativism," *Philosophical Studies* 119: 215–42. (Argues for a relativist treatment of vague terms and knowledge attributions.)
Richard, M. (2008) *When Truth Gives Out* Oxford University Press. (Argues for a relativist treatment of vague terms, and against a relativist treatment of taste predicates.)
Zimmerman, A. (2007) "Against Relativism," *Philosophical Studies* 133: 313–48. (Argues against a variety of relativist proposals.)

1.12

VAGUENESS

J. Robert G. Williams

Puzzles of Vagueness

Taking away grains from a heap of rice, at what point is there no longer a heap? It seems small changes—removing a single grain—can't make a difference to whether or not something is a heap; but big changes obviously do. How can this be, since the big changes are nothing but the small ones chained together? Call this the "little-by-little" puzzle.

At each stage, while removing grains from the original heap, ask yourself: "Is what I have at this moment a heap?" At the initial stages, the answer will clearly be "yes." At the late stages, the answer will clearly be "no." But at intermediate stages, the question generates perplexity: It's not *clearly* right to say "yes," nor is it *clearly* right to say "no."A hedged response seems better: "It's a borderline case of a heap"—but that's not yet a *direct* answer. So what is the answer to the original question, when confronted with what we can all agree to be a borderline case of a heap? Call this the "borderlineness" puzzle.

The little by little puzzle leads to the *sorites paradox* (from "sorites"—the Greek word for "heap"). Here's one version. Take a line of 10,001 adjacent men, the first with no hair, the last with 10,000 hairs, with each successive man differing from the previous by the addition of a single hair (call this a *sorites series* for "bald"). "Man N" names the man with N hairs.

Man 0 is bald. Man 10,000 is not bald. Now consider the following collection of horrible-sounding claims:

(1): Man 0 is bald, and man 1 is not bald.
(2): Man 1 is bald, and man 2 is not bald.
. . .
(10,000): Man 99,999 is bald, and man 10,000 is not bald.

If we *reject* the above, surely we must *accept* their negations:

(1*): it is not the case that: Man 0 is bald, and man 1 is not bald.
(2*): it is not the case that: Man 1 is bald, and man 2 is not bald.
. . .
(10,000*): it is not the case that: Man 99,999 is bald, and man 10,000 is not bald.

But given the various (N*), and the two obvious truths about the extreme cases, a contradiction follows. Each (N*) is (classically) equivalent to the material conditional reading of:

(N**) if Man N-1 is bald, then Man N is bald

Since Man 0 is bald, iterated *modus ponens* reasoning leads to the conclusion that Man 10,000 is bald, contrary to our assumptions. We must either find some way of rejecting seemingly compelling premises, or find a flaw in the seemingly valid reasoning.

We turn next to the puzzle of borderlineness: If Harry is intermediate between clear cases and clear noncases of baldness, "Is Harry bald?" seems to have no good, direct, answer.

There are familiar cases where we cannot answer such questions: Having never seen Jimmy, I am in no position to say whether he's bald, simply because I lack relevant knowledge. Lack of knowledge of the answer might also be appealed to in borderline cases. But for such borderline cases, appeal to ignorance simply moves the challenge one stage back. *Why* would we lack knowledge over whether Harry is bald? After all, it seems puzzlement persists even if we find out all the relevant underlying facts about the number and distribution of hairs on Harry's head. What kind of status is this, where the question of baldness is not only something we can't answer, but where we can't even conceive of *getting in a position to answer*?

It's common to say that it's *indeterminate*, or *indefinite*, or *that there's no fact of the matter* whether borderline-bald Harry is bald. Such terminology connects borderlineness to a broader class of putative examples of indeterminacy. Examples include partially defined terms, identities across theory change in science, the problem of the many, ungrounded sentences that generate semantic paradoxes, and indeterminate conditionals. It is a vexed question whether a unified analysis can be given of some or all of these "indeterminacy" phenomena, and what shape it should take if so. (Note too that while a general theory of indeterminacy might answer some questions about the status of borderline cases, it's not at all clear whether it would say anything interesting about the little-by-little puzzle.)

Sharp Extensions

Sometimes it is suggested that vague language poses obstacles to the very *idea* of providing a classical (model-theoretic) semantics for natural language. Is this right?

Focus, for simplicity's sake, on the basic case of extensional semantics. This assigns *extensions* (sets of objects) as the semantic values of predicates. This may be thought to be problematic when the predicates are vague. Sets of objects, as traditionally conceived, are definite totalities—each object is either definitely a member, or definitely not a member. Wouldn't associating one of these totalities with "bald" force us, unjustifiably, to "draw sharp boundaries" for its application? On the other hand, it seems that we *easily* say which set should be the extension of "is bald." It should be {x: x is bald}, i.e. the set of things which are bald. We can *use* vague language ("in the theorist's metalanguage") to *say* what the extension should be.

Delicate questions about methodology in semantics arise. Can semantic theory say simply that the extension of "red" (in English, or "rouge" in French) is *the set of red things*? Or should we require it give a more illuminating characterization of that set?

Even on the latter view, so long as the illuminating characterization is given in vague terms, there's no obvious obstacle to specifying the extension of a vague term. Thus Lewis, in discussing his semantic treatment of *counterfactuals* in terms of the admittedly vague notion of similarity, says "I . . . seek to rest an unfixed distinction upon a swaying foundation, claiming that the two sway together rather than independently" (Lewis, 1973, p. 92). It would be problematic if one imposed the requirement that we should be able to give a semantics for a vague object language within a *nonvague* metalanguage. But such a requirement would need to be argued for.

Perhaps, though, the worry is that terms such as {x: x is red} picks out a "vague set"—an entity not covered by standard set theories. Consider, for example, the set of bald men B, and the set B+ that is the union of B with {Harry}, where Harry is borderline bald. Since it's indefinite whether Harry is bald, it seems it will be indefinite whether Harry is a member of B; but he's definitely a member of B+. All other individuals will be a member of B iff they are a member of B+. It follows that it is indefinite whether B and B+ contain exactly the same members. By the axiom of extensionality, two sets are identical iff they have the same members. So it seems as if it will be indefinite whether B and B+ are identical.

This may be problematic. Indefiniteness in identity is the target of the famous Evans argument (Evans 1978, Salmon 1981). This purports to show, via an appeal to Leibniz's law, that such claims can be reduced to absurdity: Facts about identity are never vague. So, not only do we require an account of vague sets in general, we also need to find some way of addressing the Evans–Salmon argument.

However, a supporter of textbook (classical) semantics has responses. One idea is that terms such as "{x: x is red}" are *vague* descriptions of *precise* sets. Suppose the relevant domain D contains a single object a, and it is indefinite whether a is red. The candidates to be picked out by "{x in D: x is red}" are the null set and the singleton {a}. It might be indefinite which of the two it picks out—but a clear mistake to think it picks out some special entity, a "vague set," indefinitely identical to both. Indeterminacy in reference is a well-known escape route from the Evans reductio argument (cf. Lewis 1988). If this line succeeds, no retreat from classical conception of sharply bounded sets would be needed in order to appeal to underpin their use in semantics.

However, even if textbook semantics were fine, it wouldn't *explain* the two puzzles of the previous section: borderlineness and the sorites. In the next three sections I present examples of three kinds of positive accounts of vagueness and how they tackle the puzzles identified above. I will start with the one that requires least modification of standard logic and semantics—epistemicism. Each is successively more revisionary.

Classical Vagueness: Epistemicism

If we stick with classical logic and semantics for vague language, a consequence is that room for maneuver in explaining the puzzles of borderlineness and the sorites paradox is drastically reduced. Classicists cannot, for example, say that borderline-bald Harry is special because it is neither true nor false that he is bald—for this would conflict with the classical commitment to *bivalence* (every claim being either true or false). And unless we want to deny that Jerry Garcia in his prime was bald, and Kojak was not, then it follows, classically, that in the baldness sorites series, the last nonbald man differs by only a hair from a man who is bald.

The last claim seems particularly unpalatable—it seems to commit us, incredibly, to sharp boundaries to baldness. Some endorsing the classical logic/semantics package argue that no such commitment follows (McGee and McLaughlin 1994), but the theorists I discuss in this section take the hard-nosed stance that there really are sharp boundaries for vague predicates. To make a case for this surprising claim, at a minimum we need some explanation of how the characteristic features of vagueness are consistent with sharp boundaries.

Epistemicists typically endorse the face-value classicism just described. It's either true, or false, that borderline-bald Harry is bald, for example—there is a fact of the matter—but we do not and (in a certain sense) cannot know which option is realized. Below, I sketch Williamson's (1994) elaboration of this idea.

Let us consider borderlineness first. Start from the idea that we are ignorant of whether Harry is bald, when he is a borderline case. The puzzle was to explain why this was so, and why the unknowability was of such a strong and ineliminable sort.

Williamson's proposal makes use of a general constraint on knowledge: the idea that in order to know that p, it cannot be a matter of luck that one's belief is that p is true. Williamson articulates this as a "safety principle," roughly:

> (SAFETY)
> For "S knows that p" to be true (in such situation s), "p" must be true in any marginally different situation s* (where one forms the same beliefs using the same methods) in which "S believes p" is true.

The idea is that the situations s* represent "easy possibilities": falsity at an easy possibility makes a true belief too lucky to count as knowledge.

This first element of Williamson's view is independently motivated epistemology. The second element is that the extensions of vague predicates, though *sharp*, are *unstable*. They depend on exact details of the patterns of use of vague predicates, and small shifts in the latter can induce small shifts in the (sharp) boundaries of vague predicates.

Given these two, we can explain our ignorance in borderline cases. A borderline case of "bald" is just one where the boundary of "bald" is close enough that a marginally different pattern of usage could induce a switch from (say) Harry being a member of the extension of "bald" to his not being in the extension of that predicate. If that's the case, then even if one truly believed that Harry was bald, there will be an easy possibility where one forms the same beliefs for the same reasons, but that sentence is false. Applying (SAFETY), the belief won't count as knowledge.

Given that the source of ignorance resides in the sharp but unstable boundaries of vague predicates, one can see why gathering information about hair distributions won't overcome the relevant obstacle to knowledge. This is why the ignorance in borderline cases seems ineliminable.

What does the epistemicist say about the sorites? Whether we present that argument via a string of negated conjunctions, or conditionals, or variations and generalizations thereof, one of the premises will be false: A certain man will be bald, while his neighbor with just one more hair is not bald. The sorites argument is therefore unsound. But it's controversial whether this is enough to *resolve* our initial little-by-little puzzle. We'd like to know why we found the idea of a sharp cutoff so incredible (especially since there is available a very simple, valid argument from obvious premises to this effect). Williamson can give an account of why we'd never feel able to accept—since we

cannever know—any one of the individual conjunctions. But that doesn't explain why we're uneasy (to say the least) with the thought that *some such conjunction is true*. An analogy: I'll never know in advance which ticket will win a lottery; but I'm entirely comfortable with the thought that *one will win*. Why don't we feel the same about the sorites?

Supervaluationism

A common reaction to borderline-bald Harry is that it's *neither true nor false* that he is bald. Assuming that one can only know what is true, this would explain our inevitable lack of knowledge in borderline cases. It's often thought to be a rather plausible suggestion in itself. Classical semantics builds in the principle that each meaningful claim is either true or false. So if we're to pursue the thought that borderline claims are truth-value gaps, we must revise our semantic framework to some extent. Indeed, we can know in advance that any semantic theory with truth-value gaps will diverge from classical semantics even on some of the most intuitively plausible consequences: For it can be shown under very weak assumptions that truth-value gaps are incompatible with accepting disquotational principles such as: "Harry is bald" is true if and only if Harry is bald (see Williamson 1994, ch. 7).

How will the alteration of the classical framework go? One suggestion goes under the heading "supervaluationism" (see inter alia Fine 1975, Keefe 2000). As we'll see, the term is somewhat ambiguous. As an account of the *nature* of vagueness, supervaluationism is a view on which borderlineness arises from what we might call "semantic indecision" (cf. Lewis 1993). Think of the sorts of things that might fix the meanings of words: conventions to apply the word "bald" to clear cases; conventions to apply "not bald" to clear noncases; various conventions of a more complex sort—for example, that anyone with less hair than a bald person should count as bald. The idea is that when we list these and other principles constraining correct interpretation, we'll be able to narrow down the space of acceptable (and entirely classical) interpretations of English—but not to the single *intended* interpretation hypothesized by classical semantics. At best, what we'll get is a cluster of candidates. Let's call these *the sharpenings* for English (sometimes these are called "precisifications" or "delineations"). Each will assign to each vague predicate a sharp boundary. But very plausibly the location of such a boundary is something the different sharpenings will disagree about. A sentence is indeterminate (and if it involves a vague predicate, is a borderline case) just in case there's a sharpening on which it comes out true, and another on which it comes out false.

As an account of the *semantics* of vague language, the core of the supervaluationist proposal is a generalization of the idea found in classical semantics, that for something to be true is for it to be true *at the intended interpretation*. Supervaluationism offers a replacement. It works with a set of "co-intended interpretations," and says that for a sentence to be true, it must be true at *all the co-intended interpretations* (this is sometimes called "supertruth"). This dovetails nicely with the semantic indecision picture, since we can take the "co-intended interpretations" to be what we called above the *sharpenings*. When a sentence is indeterminate (true on one sharpening and false on another) neither it nor its negation will be true; and hence we have a *truth-value gap*.

The slight tweak to the classical picture leaves a lot unchanged. Consider the tautologies of classical logic, for example. Every *classical* interpretation will make them true; and so each sharpening is guaranteed to make them true. Hence, any classical tautology is always supertrue, which is enough to make it a *supervaluational* tautology. (It's a

matter of dispute whether more subtle departures from classical logic are required, and whether this matters. (Cf. Fine 1975, Williamson 1994 ch 5, Williams 2008.)

If (super)truth is a constraint on knowledge, supervaluationists can explain why we can't know whether borderline bald Harry is bald. On some developments of the position, they can go interestingly beyond this. One might argue that insofar as one should only invest credence in a claim to the extent one believes it true, obvious truth-value-gaps are cases where we should *utterly* reject (invest no credence in) both the claim and its negation (cf. Field 2003). This would mean the information that such-and-such is borderline gives us a direct fix on what our *degree of belief* should be in borderline cases.

What about the sorites? No individual conjunction "Man N is bald and man N + 1 is not bald" will be true—many of them will instead be *truth-value gaps*, true on some sharpenings and false on others. On the other hand, *every* sharpening can provide a number n that makes "for some n, man n is bald and man n+1 is not bald" true—different ones on different sharpenings—so it's supertrue overall. This highlights one of the distinctive (and perhaps disturbing) features of supervaluationism—that disjunctions and existential generalizations can be true, even if no disjunct or instance is. Either way, not all the premises of the premises of the argument for paradox will be true, so the argument is blocked. Just as for the epistemicist, there is a *further* question about what accounts for the argument's original appeal.

Two points to bear in mind about supervaluationism. First, as we've seen, the supervaluationist *endorses* the claim that "there is a cutoff for bald"—a pair of men differing by only one hair, with the first bald and the second not. The supervaluationist must try to persuade you that once you understand the sense in which *there is no fact of the matter* where that cutoff is, you will be prepared to accept the existence of cutoffs. Second, many want to press the charge that the supervaluationist makes no progress over the classicist, for reasons of "higher order vagueness." The thought is that the task of explaining how a *set of sharpenings* gets selected by the meaning-fixing-facts is no easier or harder than explaining how a *single classical interpretation* gets picked out. However, (a) the supervaluationist can reasonably argue that if she spells out the notion of "sharpening" in a vague metalanguage, she will regard the boundary between the sharpenings and nonsharpenings as vague (see Keefe 2000); (b) *even if* both epistemicist and supervaluationist were both in some sense "committed to sharp boundaries," the account they give of the nature of vagueness is vastly different, and we can evaluate their positive views on their own merits.

Many-Valued Settings

A distinctive feature of supervaluationism was that while it threw out *bivalence* ("Harry is bald" is either true or false) it preserved the corresponding instance of *excluded middle* ("Harry is bald or Harry is not bald"). Revising logic and semantics in a more thoroughgoing way would allow for a coherent picture where we can finally reject the claim "there is a single hair that makes the difference between bald and nonbald" without falling into paradox.

"Many valued" logics can be characterized by *increasing the number of truth-values we work with*—perhaps to three, perhaps infinitely many—and offering generalizations of the familiar stories of how logical constants behave to accommodate this tweak. There are many ways to develop this (for a survey, see "Many-Valued Logic," this volume).

Here is a sample many-valued logic, for a propositional language with conjunctions, disjunctions, and negations. To characterize the logic, we postulate three values: call them, neutrally, "1" ,"0.5," and "0." For the propositional case, each atomic sentence will be assigned one of these truth-values. The truth-values then get assigned to complex sentences recursively. Let us assume a conjunction will be assigned the minimum of the truth-values of its conjuncts; a disjunction will get the maximum of the truth-values of its disjuncts; and a negation will be 1 minus the truth-value of the claim negated (you can easily check that, ignoring the value 0.5 altogether we get back classical truth-tables).

One option at this point is to *take this model theory seriously*—much as the classicist and supervaluationist do—and hypothesize that natural language has (or is modeled by) some many-valued interpretation (or set of interpretations). The most famous version of this many-valued approach, the "degree theory" of vagueness, works not with the three values above but with *infinitely many*, labeled by the real numbers between 1 and 0. The truth-values are thought of as "degrees of truth"—with a sentence getting closer to "perfect truth" insofar as its truth-value approaches 1. (Cf. inter alia Machina 1976, Smith 2008.)

Degree theorists will say that a borderline case of a vague predicate has an intermediate truth-value. *Harry is bald* is neither wholly true, nor wholly false—and this is the reason why we are uncomfortable giving either answer. Along a sorites series, the degree of truth of "Man N is bald" falls off smoothly: for N = 0 this may be value 1; for large N it is a value close to 0. What one says about the sorites argument itself—the truth-values of the compound claims involved, and the validity of the argument, will depend on exactly what story is given about the behavior of the connectives and the characterization of the logic.

Worries about "higher order" vagueness are more pressing when directed against the degree theorist rather than against the supervaluationist. On the degree-theoretic story, it appears that Harry will be bald to some particular, precise degree—say 0.5456. But is it really plausible that there are such sharp facts of the matter about the *exact degree* to which Harry is bald? The "higher order indeterminacy" claim is that it is *indeterminate* to what degree Harry is bald. But if this higher-order indeterminacy is handled in a many-valued manner, then the language in which we ascribe degrees of truth to sentences—and hence theorize about degree theory itself—is non-classical. It's then not clear that the kind of claims we made earlier to characterize the position, e.g. "every atomic sentence has a degree of truth," are fully true. Many-valued model theory is usually described in a metalanguage tacitly assumed to be classical. It's far from clear this is dispensable (cf. Williamson 1994 ch. 4).

Some claim many-valued theories give counterintuitive results. Take a 0.5-valued sentence A. From the recursive clauses, we can see that "A&~A" will be half-true as well. Many find this commitment objectionable. (See Williamson 1994 ch. 4 for arguments that *all* many-valued treatments of the connectives will be similarly problematic.) Degree-theorists don't agree, and standoff threatens. It's worth noting that this objection relies on a particular "degrees of truth" gloss on the many-valued semantics. But in a three-valued setting, rather than talk of degrees of truth, one could interpret value 1 as "true," 0 as "false," and 0.5 as "neither true nor false" (cf. Parsons 2000). On this reading, "A&~A" will be described as a truth-value gap, when "A" itself has this status—which sounds rather different (and more plausible) than describing it as half true. The opponent of many-valued approaches must check that their objections don't depend on a contentious gloss on the targeted view.

We've concentrated till now on accounts that *take model theory seriously*—and which look for an integrated account of the logic and semantics of vague language. But there's another option—to focus attention on the logic, and regard model theory as simply a descriptive instrument we use to specify the One True Logic.

Logic-first theorists advocate a revisionary take on what *arguments are valid* and what *combinations of claims or attitudes are consistent*. To fix ideas let's focus on an example, using the 3-valued assignments described earlier. Recall that a sequent is valid on the *strong Kleene logic* if it is "1-preserving" across these assignments, i.e. if it is such that when all the premises are value 1, the conclusion is value 1 too. In this Kleene logic, "Av~A" is no longer a tautology, since it need not be value 1. On the other hand, "A&~A" is still treated as a logical contradiction (every sentence whatsoever follows from it), since it will never attain value 1, no matter what value A is assigned.

For the Assignment-first theorists we have been considering thus far, the interesting question was about the various assignments and what they represent. Does the middle status represent a truth-value gap or a degree of truth? Among the variety of assignments compatible with the rules, which matches up to the actual distribution of truth statuses? But for a logic-first theorist, such questions are misguided. The assignments are simply mappings from sentences to numbers—there's no need to think that they latch on to real properties of the sentences, and so no sense in asking about what they represent or which one among them is "intended"—all they do is collectively characterize a Kleene-style logic.

A logic-first theorist such as Field (2003a) then uses the logic to handle the distinctive puzzles of vagueness. Field is committed to a certain *normative significance* for logic—that accepting the premise of a valid argument commits one to accepting its conclusion; and rejecting the conclusion commits one to rejecting its premise. Note that the Kleene logic does not make excluded middle a tautology, so on Field's view there's no *logical* reason to accept instances of it. Field advocates *rejecting* the claim that "Harry is bald or he isn't" when Harry is borderline-bald. Since this disjunction follows, in the logic he endorses, from "Harry is bald," he rejects this claim too, as well as its negation. This then is Field's account of the puzzle of borderlineness: Both direct answers to "Is Harry bald?" are to be rejected.

Note that this story would collapse if to reject a claim is simply to accept its negation—we must be able to reject "Harry is bald" without accepting "Harry is not bald." That situation can get more dramatic. In the Kleene logic, the negation of excluded middle "~(Av~A)" is inconsistent (and so must never be accepted). But we are supposed to reject instances of Av~A. So a basic feature of the position is that sometimes we reject claims whose negations are inconsistent.

One can view the sorites paradox as repeating this pattern on a grander scale. Instances of "Man N is bald and Man N+1 isn't" do indeed sound horrible, and we should reject them. But the sorites paradox does not get started unless we move from this to endorsing their negations. What the sorites paradox reveals is that the negations generate a contradiction. But just as with a single instance of excluded middle, there's no obligation to move from rejecting the claim to endorsing its (problematic) negation.

Clearly, the logic-first version of many-valuism is very different from the assignment-first version. Note, for example, that the higher-order indeterminacy worries mentioned earlier simply do not get started against the logic-first theorists (at least if they endorse the instrumental view of model theory). If one's account has no place for "the intended

truth-value assignment," then opponents can hardly ask embarrassing questions about whether it's determinate which assignment this is!

On the other hand, the logic-first position is radical. It gives up on the attractive prospect of a model-theoretic *analysis* and *explanation* of validity. And it seems committed to doing without model theory when directly addressing questions about the semantics and truth status of vague language (Field, for example, endorses a nonsemantic, *disquotational* treatment of truth and meaning).

Conclusion

I've outlined three sample positive views of vagueness. They are illustrative of broad *categories* of theories of vagueness. The categories are not exhaustive, but they do orientate much contemporary debate.

The most conservative category–represented here by epistemicism–preserves classical logic, model theory and textbook semantics. The distinctive *epistemicist* response is one way of explaining what vagueness could be, even if the apparent commitments of this setting to "sharp cutoffs" are taken at face value. But this isn't the only way a classicist might respond. Fara (2003), for example, agrees with Williamson on the *broad classical setting*. But her account of what vagueness is—at least for the cases she focuses on, gradable adjectives—is very different. Her idea is that (i) the word "red" means something like *significantly redder than is typical*; (ii) what counts as "significant" is deeply interest-relative, and in particular, can vary depending on the focus of our attention; (iii) as a result of this, we can predict that whenever we focus on a particular potential "cutoff" for red, the facts about what is now *significant* to us ensure that the classical cutoff for "red" is located somewhere *other* than we are looking. Hence, for Fara, the distinctive elusiveness of vague predicates. (The idea that context-sensitivity might be intimately involved with the phenomena of vagueness is something many writers find appealing, whether or not they work in a classical backdrop (see Raffman 1994, Shapiro 2006).)

The next most conservative response is to design a "special purpose" semantics for vagueness, allowing semantic features to do significant explanatory work, which nevertheless "saves much of the appearance" of the standard classical setting. Supervaluationism illustrates this possibility. But there are other ways to go, even with the broad "semantic indecision" framework that gives us our notion of a *sharpening*. Various authors have suggested that we might look at the *proportion* of sharpenings on which a given claim holds. In terms of this, we can characterize a notion of *intermediate degrees of truth*: the degree of truth of "Harry is bald" being the proportion of sharpenings on which it is true. Thus supertrue sentences ("Kojak is bald") will be true to degree 1; while if "Harry is bald" is plumb borderline, true on half and false on half, then it'll be true to degree 0.5 (cf. Edgington 1997). Such notions may find application within a treatment of comparative forms of vague adjectives and modifiers (see Kamp 1975, Lewis 1970).

The final theoretical framework we looked at was thoroughly revisionary, though the revisionism has more or less radical forms. Preserving realism about model-theoretic foundations, we find "gap" and "degree" theories. Alternatively, looking at model theory as a mere descriptive instrument for specifying a logic, we have logic-first theories. There's plenty of room for debate about which logic is most appropriate: For example, Crispin Wright (2003, 2007) argues for *intuitionism* as the appropriate logic governing vague discourse. (Wright has been a long-time advocate of drawing lessons

on the nature of semantic theory from careful study of the phenomenon of vagueness–see Wright (2007).)

Important though these logico-semantic issues are, one should not think that the whole task of understanding vagueness consists in figuring out the proper semantic treatment of it. As is evident from the case of epistemicism, resources outside the philosophy of language may be key to diagnosing its characteristic puzzles. Moreover, logico-semantics may *underdetermine* the overall theory of vagueness. Getting clear on the *nature* of vagueness/indeterminacy is another aspect to theorizing about vagueness –for example is it a matter of "semantic indecision" as many supervaluationists contend? Similar formal machinery can be paired with very different responses to this question. As well as the *nature, psychological* features demands attention. Should we *have zero credence* in Harry being bald, if we know he's a borderline case (as Field argues)? Or should have fifty-fifty confidence, as some degree theorists maintain? (Cf. Smith 2008, Schiffer 2003.) Or something else entirely (Wright 2003)?

The attention lavished on vagueness since the mid-1970s shows little signs of converging on a single standard account. But it has given us plenty of options, providing rich resources for future arguments. To study it is to appreciate the far-reaching impact of the disarmingly simple puzzles of vagueness.

Bibliography

D. Edgington. (1997) "Vagueness by degrees," in *Vagueness: A Reader*, eds. Rosanna Keefe and Peter Smith (Cambridge, MA: MIT Press), 294–316.

G. Evans. (1978) "Can there be vague objects?" *Analysis* 38, no. 13: 208.

D. Graff Fara. (2003) "Shifting sands: an interest-relative theory of vagueness," *Philosophical Topics* 28, no. 1: 45–81.

H. H. Field. (2003) "Semantic paradoxes and the paradoxes of vagueness," in *Liars and Heaps: New essays on paradox*, ed. J. C. Beall (Oxford: Oxford University Press), 262–311.

H. H. Field. (2003a) "No fact of the matter," *Australasian Journal of Philosophy* 81, no. 4: 457–480.

K. Fine. (1975) "Vagueness, truth and logic," *Synthese* 30: 265–300.

J. A. W. Kamp. (1975) "Two theories about adjectives," in *Formal Semantics of Natural Language*, ed. E. Keenan (Cambridge: Cambridge University Press), 123–55.

R. Keefe. (2000) *Theories of Vagueness* (Cambridge: Cambridge University Press).

D. K. Lewis. (1970) "General semantics," *Synthese* 22: 18–67.

—— (1973) *Counterfactuals* (Oxford: Blackwell).

—— (1988) "Vague Identity: Evans misunderstood," *Analysis* 48: 128–30.

—— (1993) "Many, but almost one," in *Ontology, Causality and Mind: Essays on the philosophy of D. M. Armstrong*, ed. Keith Campbell, John Bacon, and Lloyd Reinhardt (Cambridge: Cambridge University Press).

K. F. Machina. (1976) "Truth, belief and vagueness," *Journal of Philosophical Logic* 5: 47–78.

Vann McGee and Brian McLaughlin. (1994) "Distinctions without a difference," *Southern Journal of Philosophy* supp XXXII: 203–251.

T. Parsons. (2000) *Indeterminate Identity: Metaphysics and semantics*. New York: Oxford University Press.

D. Raffman. (1994) "Vagueness without Paradox," *Philosophical Review* 103.

N. U Salmon. (1981) *Reference and Essence* (Princeton: Princeton University Press).

S. Schiffer. (2003) *The Things We Mean* (New York: Oxford University Press).

N. J. J. Smith. (2008) *Vagueness and Degrees of Truth* (Oxford: Oxford University Press).

J. R. G. Williams. (2008) "Supervaluationism and logical revisionism," *The Journal of Philosophy* CV, no. 4.

T. Williamson. (1994) *Vagueness* (London: Routledge).

C. Wright. (2003) "'Vagueness: A fifth column approach'," in *Liars and Heaps: New essays on paradox*, ed. J. C. Beall (Oxford: Oxford University Press), 84–105.

—— (2007) "Wang's paradox" in *The Philosophy of Michael Dummett*, eds. Randall E. Auxier and Lewis Edwin Hahn (Peterborough, NH: Open Court).

1.13
EMPTY NAMES

Sarah Sawyer

1 Introduction

An empty name is a name that fails to name an individual—a name that fails to refer. Fiction, myth, and false theory provide plausible examples. Thus, on the face of it, 'Sherlock Holmes,' 'Pegasus,' and 'Vulcan' fail to refer precisely because there is no such detective as Sherlock Holmes, there is no winged horse Pegasus, and there is no planet Vulcan whose orbit lies between Mercury and the Sun. Nonetheless, some sentences containing empty names can be understood, and hence appear to be meaningful—for example, 'Pegasus was sired by Poseidon' and 'Vulcan was postulated by Le Verrier'; some empty names appear to name the same individual and hence to be co-referential—for example, 'Santa Claus' and 'Father Christmas'; and some sentences containing empty names appear to be straightforwardly true—for example, 'Sherlock Holmes is a character from fiction,' and 'Pegasus does not exist.' The expression 'straightforwardly true' is intended to contrast 'true according to the fiction/myth/failed theory,' which require independent treatment. Thus 'Sherlock Holmes played the violin' is, on the face of it, true according to the fiction but not straightforwardly true, whereas 'Sherlock Holmes is a character from fiction' is, on the face of it, straightforwardly true but not true according to the fiction. A semantic theory of names must make sense of the apparent phenomena of meaning, reference, and truth in the specific context of sentences containing names that fail to refer (either by accommodating the relevant intuitions or by explaining them away).

Of course, the question of which names, if any, are empty cannot be answered by philosophy of language alone. Rather, a semantic theory of names must be provided without prejudging the answer to that question. A semantic theory of names should deliver a specification of the conditions under which a name names an individual, and hence a specification of the conditions under which a name is empty; but the question of whether those conditions are met on any given occasion of use plausibly lies outside the scope of a semantic theory. Having said that, there are those who have attempted to make sense of meaning, reference and truth in the context of *prima facie* empty names by postulating referents for those names. This strategy aims to eliminate the problems surrounding empty names by eradicating (or reducing to an insignificant minimum) the category of empty names. Terence Parsons has defended the claim that the referents of apparently empty names are nonexistents (Parsons 1980); Nathan Salmon and Amie Thomasson have argued that they exist as abstracta (Salmon 1987; Salmon 1998; Salmon 2002; Thomasson 1998). In what follows I will provide an overview of semantic

theories which address the need to make sense of the apparent phenomena of meaning, reference, and truth in the specific context of sentences containing empty names.

2 Fregean Sense

One may wonder why exactly empty names are thought to be problematic. The problem is strikingly clear, however, if set against a certain intuitive understanding of how names and predicates function. Intuitively, the function of a name is to pick out (refer to) an individual, and the function of a predicate is to pick out (refer to) a property. If this is right, it is natural to think: first, that a sentence composed of a name and a predicate is true just in case the object referred to by the name has the property referred to by the predicate (which presupposes that there is an object named); and, second, that a subject can understand a sentence composed of a name and a predicate only if she knows which object is referred to by the name and which property is referred to by the predicate (which, again, presupposes that there is an object named). If this basic understanding of the function of names and predicates is correct, empty names are essentially defective, and we are bound to conclude: first, that sentences containing empty names can be neither true nor false (since there is no object which has or lacks the property referred to by the predicate); and second, that sentences containing empty names cannot be understood (since one cannot know which object is referred to by a name that refers to no object).

Frege's move from an extensional semantic theory akin to the one described above (a theory that countenances only linguistic expressions and their referents) to an intensional semantic theory that invokes in addition a notion of sense, embodies a rejection of the intuitive understanding of how names and predicates function. According to Frege, every linguistic expression expresses a sense, which is objective and is grasped by anyone who understands the expression. The sense of an expression determines the referent of that expression, but, crucially, the sense of an expression is available to be thought of independently of whether the expression in fact refers. Consequently, Frege's theory provides an account of how sentences that contain empty names can have meaning and be understood. In order to understand a sentence composed of a name and a predicate, one need simply grasp the sense of the sentence (the thought expressed), and this will be available even in the absence of a referent for the name. (It is typically assumed that the Fregean sense of a name must be descriptive. However, I think there is good reason to think of the sense of a name as nondescriptive. For one thing, the claim that names have descriptive senses conflicts with Frege's claim that sense is transparent: It is, after all, possible for a subject rationally to assent to a sentence composed of a name and a predicate, and simultaneously withhold assent from the sentence had by replacing that name with any description that allegedly expresses the same sense.)

The problems of apparent co-reference and of truth, however, remain. The fact that 'Santa Claus' and 'Father Christmas' appear to be co-referential cannot be explained by appeal to the senses of the expressions, since the senses of the expressions will differ. And sentences such as 'Pegasus does not exist,' which contain a name that fails to refer, will be neither true nor false, since, according to Frege, the truth-value of a sentence is its referent, and the referent of a complex expression is determined by the referents of its parts. Thus, if a part of a complex expression lacks a referent, so does the whole, which in the case of a sentence means that the sentence has no truth-value.

Frege was not primarily concerned with ordinary language. He remarks that it is a fault of languages that they contain expressions that fail to refer, and maintains that in

a logically perfect language no name would be introduced without being provided with a referent (Frege 1892: 69–70). However, the problems raised by empty names are problems that must be dealt with satisfactorily if an adequate semantics for ordinary language is to be found. As such, either Frege's remarks must be supplemented so as to account for our intuitions about reference and truth, or we must look to a different semantic theory for an answer.

Gareth Evans and John McDowell advocate a Fregean theory according to which the sense of a singular term, such as a name or a demonstrative, is object-dependent, or *de re* (Evans 1982; McDowell 1977; McDowell 1984). On this view, expressions that lack referents thereby lack sense: Consequently, a sentence containing an empty name fails to express a thought. Rather than helping matters, this appears to leave all three problems surrounding empty names unresolved—those of meaning, reference, and truth.

3 Russellian Descriptions

Russell became suspicious of intensional notions such as Fregean sense (Russell 1905), and sought instead to solve the problems raised by empty names within a semantic theory that once again countenanced only linguistic expressions and their referents. Nevertheless, Russell also rejected the intuitive understanding of how names function. Thus, according to Russell, names, such as 'Bertrand Russell' and 'Vulcan,' do not express a Fregean sense, nor do they function to pick out an object directly. Rather, according to Russell, names are disguised definite descriptions of the form 'the F.' Thus, for example, the name 'Bertrand Russell' is taken to be semantically equivalent to a definite description such as 'the male philosopher who was born on May 18,1872 and wrote "On Denoting" '; similarly, the name 'Vulcan' is taken to be semantically equivalent to a definite description such as 'the planet whose orbit lies between Mercury and the Sun.'

Further, and crucially, Russell rejected the intuitive understanding of definite descriptions as singular terms that function to pick out an object directly (which would leave the problems concerning empty names unresolved). Instead, he argued, definite descriptions such as 'The F' are quantificational phrases which fall into the same broad semantic category as expressions such as 'Some F' and 'Every F.' On this view, the function of a name, understood as a definite description, in turn understood as a quantificational phrase, is not to contribute an object to propositions expressed by sentences containing it: rather, the function of a name is to contribute a complex of properties (Russell 1905; Russell 1911; Russell 1918). The beauty of this account lies in the fact that a name has purely extensional semantic content which is nonetheless available whether or not it is the name of an individual. A name is empty in the relevant sense if and only if there is no object that uniquely satisfies its descriptive content. Thus 'Bertrand Russell' is not an empty name because there is in fact just one male philosopher who was born on May 18, 1872 and wrote "On Denoting"; whereas 'Vulcan' is an empty name because there is no planet whose orbit lies between Mercury and the Sun.

Russell's account offers a solution not only to the problem of meaning (and hence understanding) for sentences containing empty names but also to the problem of truth. Thus, taking the relevant names to be equivalent to the definite descriptions suggested above, the sentence 'Bertrand Russell was a pacifist' is to be analyzed as 'There is exactly one male philosopher who was born on May 18, 1872 and wrote "On Denoting" and he was a pacifist,' which is true in virtue of the fact that there is such an individual and he was a pacifist. In like fashion, the sentence 'Vulcan was discovered by Le Verrier' is to

be analyzed as 'There is exactly one planet whose orbit lies between Mercury and the Sun and it was discovered by Le Verrier,' which is false in virtue of the fact that there is no such planet. Finally, the sentence 'Vulcan does not exist' is to be analyzed as 'It is not the case that there is exactly one planet whose orbit lies between Mercury and the Sun,' which, in accord with intuition, is straightforwardly true.

The proposed solution to the problem of truth is not unproblematic, however. Russell's account implies that all sentences composed of an empty name and a predicate are false, including: 'Pegasus is a mythical creature' and 'Venus was incorrectly postulated by Le Verrier.' Russell was concerned to preserve the law of excluded middle, according to which every sentence is either true or false. However, it is not obvious that the law of excluded middle ought to be preserved, and the particular way in which Russell's account preserves it leaves us with unintuitive assignments of truth-values, as indicated. Moreover, once again the problem of apparent co-reference remains.

Russell's account of definite descriptions as quantificational phrases has gained a reasonable degree of support. (Although see Strawson 1950; Donnellan 1966. For a defence of Russell's account of definite descriptions see Neale 1990.) His claim that names are disguised definite descriptions, in contrast, has been largely discredited by Kripke's arguments to the effect that names and definite descriptions have different modal profiles (Kripke 1980). For instance, it is false that Bertrand Russell might not have been Bertrand Russell (although he might of course have had a different name); but it is true that Bertrand Russell might not have been the male philosopher who was born on May 18, 1872 and wrote "On Denoting" (since Russell might not have become a philosopher). As a concession to the facts about the different modal profiles of names and definite descriptions, it has been suggested that a Russellian account of names might nonetheless be advocated specifically for names that are empty. The suggestion here is that names that name an individual be treated as directly referring expressions that pick out an individual directly, while empty names be treated as having descriptive content (e.g. Currie 1990). There is some plausibility to this idea, since (one might think) empty names are introduced by allegedly reference-fixing descriptions from which they never manage to break free. The main concern is that such an account would not provide a *unified* semantic account of proper names, since the semantics of a name would then depend on whether or not it refers.

4 Millianism

Qualms about intensional notions, together with the realization that names do not in general have descriptive content, has led to a striking consensus that names are directly referring expressions. This is effectively the intuitive account with which we started, according to which the function of a name is simply to pick out (refer to) an individual. One influential kind of direct reference theory is Millianism, according to which the meaning of a name is the object named (Salmon 1986; Soames 1987; Soames 2002). Millians often maintain that a sentence composed of a name and a predicate expresses a structured proposition which contains the object referred to by the name and the property referred to by the predicate; and that a sentence containing an *empty* name expresses a 'gappy' proposition—a proposition that has a gap where an object would be if the name were not empty. Consequently, the view runs headlong into the three problems of empty names with which we started. Different attempts to solve the problems have emerged consistent with both the claim that a name is a directly referring

expression and the claim that a sentence containing an empty name expresses a gappy proposition. I will look at three primary attempts here. I leave aside the view proposed by Salmon and Thomasson (Salmon 1998; Salmon 2002; Thomasson 1998). This view avoids the problems of empty names by treating names from fiction, myth, and failed scientific theories as referring to abstract artefacts, and hence as nonempty. I have argued against such a view elsewhere (Sawyer 2002). Salmon does maintain that in rare cases a name can be empty, but such cases are so rare as to be relatively insignificant. The view proposed by Parsons (Parsons 1980), according to which apparent empty names refer to nonexistent entities, is discussed below.

The first approach I will discuss here is developed by David Braun (Braun 1993; Braun 2005). According to Braun, an atomic sentence composed of a name and a predicate expresses a proposition which is true if and only if there is an object referred to by the name and it has the property referred to by the predicate; and propositions which are not true are false. Thus every atomic sentence containing an empty name expresses a gappy proposition which is false: for example, 'Sherlock Holmes is a detective,' 'Pegasus was sired by Poseidon,' and 'Vulcan was postulated by Le Verrier.' One apparent benefit of Braun's view is that sentences of the form 'a exists' are also to be treated as atomic sentences of subject–predicate form. As a result—and in accord with intuition—sentences such as 'Sherlock Holmes exists,' 'Pegasus exists,' and 'Vulcan exists' express false propositions, while their negations—'Sherlock Holmes does not exist,' 'Pegasus does not exist,' and 'Vulcan does not exist'—express true propositions.

Less intuitive are the following consequences of the view. First, sentences that differ only in respect of the empty name they contain express the very same proposition: Thus, 'If Pegasus exists, then there is a planet between Mercury and the Sun' and 'If Vulcan exists, then there is a planet between Mercury and the Sun' both express the same, false, gappy proposition. This goes against an intuition concerning truth—that the former is false, while the latter is not—and against two related intuitions concerning meaning—namely, that the sentences differ in meaning, and hence that the names have meaning. Second, atomic propositions and their predicate-negations express propositions both of which are false: Thus 'Pegasus has wings' and 'Pegasus does not have wings' both express false propositions, which again goes against an intuition concerning truth—this time that the sentences differ in truth-value (or, perhaps, are truth-valueless).

Braun aims to accommodate the relevant intuitions concerning meaning and truth by appeal to cognitive facts—the semantic facts alone being unable to accommodate them. According to Braun, there are different ways in which one can believe or disbelieve a proposition; and each of these ways corresponds to a different mental state, which can play a different role in one's cognitive life. Thus a single proposition, whether gappy or not, can provide the content of numerous different mental states if that proposition is believed and/or disbelieved in different ways. So how does the appeal to ways of believing a proposition accommodate the intuitions about meaning and truth? According to Braun, the (false) intuition that 'If Vulcan exists, then there is a planet between Mercury and the Sun' is true, whereas 'If Pegasus exists, then there is a planet between Mercury and the Sun' is false, is to be explained by the fact that there is one, false, gappy proposition that we both believe in a 'Vulcan-ish' way and disbelieve in a 'Pegasus-ish' way. Similarly, the (false) intuition that 'Pegasus has wings' and 'Pegasus does not have wings' differ in truth-value is to be explained by the fact that we believe the false, gappy proposition expressed by the first sentence but disbelieve the false gappy proposition expressed by the second. Intuitions about differences in meaning are also

to be explained by appeal to different ways of believing. Since we necessarily believe a proposition in some particular way, false intuitions about differences in meaning can be seen as mapped onto—and hence as being sensitive to—real differences in ways of believing.

The success of this strategy depends upon the viability of the crucial notion of *ways of believing* a proposition. Appeal to ways of believing (or something similar) is prevalent among Millians as a result of the felt need to explain how a subject can take differing cognitive attitudes toward one and the same proposition. Thus, for example, although 'Cary Grant is an actor' and 'Archie Leach is an actor' express the same proposition (since 'Cary Grant' and 'Archie Leach' are co-referential, directly referring expressions), it is possible for a rational subject to believe what the first says and not believe what the second says. For the sake of argument I assume here that this strategy is plausible in the case of names that refer. However, even if an appeal to ways of believing were legitimate in cases where the relevant names refer, cases where the relevant names are empty cannot be assumed to be analogous. After all, if the name 'Vulcan' lacks meaning, then plausibly the expression 'Vulcan-ish'—which on the face of it is a parasitic term—also lacks meaning: Consequently, there is no clear meaning to the theoretical expression 'believing in a "Vulcan-ish" way.' One might think that 'believing in a "Vulcan-ish" way' could be understood metalinguistically, as equivalent to 'believing by means of the name "Vulcan" ' , but this won't do since many different things can share a name. 'Vulcan' is the name of a hypothetical planet, a Roman god, a colleague's pet tortoise, and more besides. This casts doubt on the general strategy employed, which in turn casts doubt on the ability of Braun's account to solve the problems of meaning and truth. Moreover, even if the notion of ways of believing can be made to work, it is unclear how it would solve the problem of apparent co-reference.

The second approach has been developed by Fred Adams, Gary Fuller, and Robert Stecker (Adams et al. 1994; Adams et al. 1997; Adams and Fuller 2007. See also Taylor 2000; Soames 2002). On this view, all gappy propositions expressed by sentences that contain empty names simply lack truth-value; and intuitions concerning meaning, reference, and truth are to be explained by appeal to the descriptive information that those sentences pragmatically convey. More specifically, the introduction of an empty name into the language (whether intentionally or unintentionally) occurs within the context of a set of descriptions which are intended to provide information about an individual bearing that name. These descriptions are then pragmatically conveyed by subsequent uses of the name, although the descriptions are never part of the meaning of the name. Fictional names are associated with descriptions from the relevant fiction; mythical names are associated with descriptions from the relevant myths; and names from failed scientific theories are associated with descriptions from the relevant scientific theory.

So how does the notion of descriptive information pragmatically conveyed accommodate the intuitions about meaning, reference, and truth? According to Adams, Fuller, and Stecker, since the descriptions pragmatically conveyed by 'Vulcan' are different from those pragmatically conveyed by 'Pegasus,' it seems to us that the names differ in meaning. The names 'Santa Claus' and 'Father Christmas,' in contrast, plausibly pragmatically convey the same set of descriptions. They say: "There are two ways that a set of descriptions may come to be associated with more than one name. An author may make the associations explicit in the text. . . . Also, an association may be arrived at as an inference to the best explanation" (Adams et al. 1997: 132). Further, despite the fact that 'Pegasus exists' and 'Pegasus does not exist' both lack truth-value, according

to Adams, Fuller, and Stecker, we think the former is false and the latter is true because the former pragmatically conveys the information that a winged horse exists, which is false, whereas the latter pragmatically conveys the information that a winged horse does not exist, which is true. Our intuitions concerning meaning, reference, and truth, then, are explained by our mistaking the information a sentence *pragmatically conveys* for the proposition it *expresses*.

There are at least four problems with the view. (For the first, second, and third see Everett 2003; for the fourth see Reimer 2001.) First, the difference in modal profile between names and definite descriptions is equally evident between empty names and the sets of descriptions they pragmatically convey. Thus 'Pegasus is Pegasus' is necessarily true, whereas 'Pegasus helped Bellerophon in his fight against the Chimera' is not. Consequently, the modal profile of a sentence containing an empty name cannot be explained by appeal to the pragmatic information it conveys; nor can it be explained by appeal to the proposition expressed, which is gappy and necessarily lacking in truth-value. Second, if we are prone to mistake the information a sentence pragmatically conveys for the proposition it expresses in the case of sentences containing empty names, then we ought to be prone to the very same mistake in the case of sentences that contain names that refer, but clearly we are not. Kripke's arguments to the effect that names are not equivalent to definite descriptions depends upon us not being subject to this kind of confusion. Third, intuition favors the thought that the meaning of a sentence containing an empty name is determinate, whereas the descriptions associated with any given name will vary across speakers and times. And fourth, the relevant semantic intuitions about empty names persist even once we are offered a pragmatic explanation of them. Successful pragmatic explanations of semantic intuitions, however, tend to rid us of the initial intuitions they are invoked to explain. This indicates that the pragmatic explanation offered by Adams, Fuller, and Stecker is not after all a genuine explanation of the apparent phenomena of meaning, reference ,and truth surrounding empty names.

The fundamental flaw in the pragmatic approach is that it loses sight of the intuition that an empty name is a name: that it is a directly referring expression that allows us to talk about an object independently of the object's satisfying any descriptions associated with it. No doubt there are descriptive propositions pragmatically conveyed by a sentence containing an empty name—just as there are descriptive propositions pragmatically conveyed by a sentence containing a non-empty name—but appeal to what is pragmatically conveyed cannot do justice to our intuitions about meaning, reference, and truth.

The third, rather different, approach is advocated by Marga Reimer (Reimer 2001). According to Reimer, the intuitions concerning meaning, reference, and truth are all to be accounted for by the fact that "as speakers of the language, we unreflectively assume that there are nonexistents and that reference to them is possible" (Reimer 2001: 499). It is important to realize that Reimer's approach does not require that the assumptions we make themselves be true—it may well be incoherent to assume there are nonexistents to which we can refer. But it is, as she points out, clear that the assumptions are made. This is why we are willing to accept as meaningful and true sentences such as 'There is a fictional character called "Sherlock Holmes", but he doesn't exist—he's only fictional after all' (Reimer 2001: 499). The acceptance of a philosophical theory (such as Millianism, for example) may well disrupt our pretheoretic intuitions. But the problems surrounding empty names concern our pretheoretic intuitions, and these are to be explained by the pretheoretic assumption of reference to nonexistents.

Reimer's approach appears to explain all three apparent phenomena of meaning, reference, and truth. Essentially this is because the apparent phenomena are based on the fundamental fact that an empty name is a name. It would be tempting to conclude that this provides further support for Millianism as a semantic theory. But such a conclusion would be too hasty. What Reimer has identified is a plausible assumption that speakers make which explains our inuitions concerning meaning, reference, and truth in the context of sentences containing empty names. The assumption—together with its ability to explain the relevant intuitions—is certainly consistent with Milliansim: but the two are strictly independent. In particular, the assumption is also consistent with a view proposed by Tyler Burge according to which a name is a predicate true of an object if and only if the object was given that name in an appropriate way (Burge 1973). Names, if Burge is right, are predicates potentially true of many different individuals. According to Burge, a singular use of a name such as 'Bertrand Russell' as it occurs in sentences such as 'Bertrand Russell was a pacifist' are to be understood as semantically equivalent to demonstrative expressions such as 'that Bertrand Russell', and hence as falling within the same semantic category as explicit demonstrative-predicate constructions such as 'that cat.' An empty singular use of a name such as 'Vulcan' as it occurs in sentences such as 'Vulcan is a planet' are then treated as failed demonstratives. Since this is the only semantic theory of names that adequately accounts for the fact that different individuals can share a name (Burge 1973; Sawyer 2010), it is this semantic theory in conjunction with the assumption Reimer identifies that ought to be accepted.

5 Empty Predicates

The problems of meaning, reference, and truth that surround empty names have been widely discussed throughout the history of analytic philosophy, and the ability of a semantic theory to deal with the problems has been taken as a primary consideration in its assessment. In contrast, it is widely assumed that no such problems arise for predicates: that predicates are semantically unproblematic in virtue of being guaranteed a referent. This, I think, is misguided, particularly in a naturalistic climate. It is no more obvious that predicates such as 'x is a unicorn' and 'x is a Greek god' are guaranteed a referent, than that names such as 'Pegasus' and 'Poseidon' are guaranteed a referent.

This is potentially problematic in particular for semantic theories which try to deal with the problems surrounding empty names by appeal to the guaranteed referents of predicates. Russell's theory of descriptions provides an early example of this. According to Russell, a name would be problematic if understood as a singular term, since a singular term makes no semantic contribution to a proposition expressed by a sentence containing it if it lacks a referent; but an ordinary name is not problematic if understood as a definite description, since a definite description always provides a semantic contribution to a proposition expressed by a sentence containing it, even if there is no object which uniquely satisfies that description. But if it were possible for a definite description to be empty—not in the sense of there being no object that satisfies it, but in the sense of there being no set of properties it refers to—the problem of empty names would not have been solved.

A second example is provided by Parsons's theory of nonexistent objects (Parsons 1980). According to Parsons, there is an object corresponding to every set of properties. To some of those sets of properties there corresponds an object that exists, and to others there corresponds an object that does not exist (a nonexistent object). The realm

of nonexistent objects includes objects from fiction, myth, and failed scientific theory, as well as impossible objects such as round squares. Consequently, Parsons's ontology of nonexistents solves the semantic problems raised by empty names because there are (effectively) no longer any empty names: there are simply names that refer to existent objects on the one hand, and names that refer to nonexistent objects on the other. (Parsons does distinguish referring to a nonexistent object from failing to refer to an object, but the distinction does not affect the main point here.) But, crucially, the solution to the semantic problems depends on the claim that every predicate refers to a genuine property, and in particular that the predicates that define the nonexistent objects refer to genuine properties. If 'x is a Greek god' were an empty predicate, then there could be no Greek gods, not even nonexistent ones.

A rather different kind of theory that relies on the same questionable presupposition is the version of Millianism that is committed to the claim that sentences containing empty names express 'gappy' propositions but tries to account for our intuitions concerning meaning, reference and truth by appeal to the descriptive content either of the sentences themselves, or of the propositions pragmatically conveyed by them. Again, the presupposition here is that the availability of descriptive content provided by the predicates is unproblematic (Recanati 1993; Adams et al. 1994; Taylor 2000; Adams and Fuller 2007). As discussed above, Braun's appeal to ways of believing also relies on predicates being guaranteed a referent, since 'x believes in a "Vulcan-ish" way' must refer if his strategy is to work.

If predicates are guaranteed referents, then appealing to the referents of predicates in an attempt to solve the problems of meaning, reference, and truth for empty names might be legitimate. If not, then alternative solutions must be found, ones that can in addition solve the parallel problems of meaning, reference, and truth for empty predicates. Reimer's claim that our pretheoretic use of empty names involves the presupposition that we can refer to nonexistents could perhaps be adapted here; in like fashion, our pretheoretic use of empty predicates involves the presupposition that they refer to properties. The relevant presuppositions, however, are consistent with more than one semantic theory, and should not themselves be seen as part of any given semantic theory.

References

Adams, F., Fuller, G., & Stecker, R. (1994) 'Vacuous Singular Terms,' *Mind and Language* 9: 387–401.

Adams, F., Fuller, G., & Stecker, R. (1997) 'The Semantics of Fictional Names,' *Pacific Philosophical Quarterly* 78: 128–48.

Adams, F. and Fuller, G. (2007) 'Empty Names and Pragmatic Implicatures,' *Canadian Journal of Philosophy* 37: 449–62.

Braun, D. (1993) 'Empty Names,' *Noûs* 27: 449–69.

Braun, D. (2005) 'Empty Names, Fictional Names, Mythical Names,' *Noûs* 39: 596–631.

Burge, T. (1973) 'Reference and Proper Names,' *Journal of Philosophy* 70: 425–39.

Currie, G. (1990) *The Nature of Fiction*, Cambridge: Cambridge University Press.

Donnellan, K. (1966) 'Reference and Definite Descriptions,' *Philosophical Review* 75: 281–304.

Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press.

Everett, A. (2000) 'Referentialism and Empty Names,' in Everett and Hofweber (eds.) *Empty Names, Fiction and the Puzzles of Nonexistence*, Stanford: CSLI Press.

Everett, A. (2003) 'Empty Names and "Gappy" Propositions' *Philosophical Studies* 116: 1–36.

Frege, G. (1892) 'On Sense and Meaning,' translated in P. Geach and M. Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1952.

Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.

McDowell, J. (1977) 'On the Sense and Reference of a Proper Name,' *Mind* 86: 159–85.
McDowell, J. (1984) '*De Re* Senses,' *Philosophical Quarterly* 34: 283–94.
Neale, S. (1990) *Descriptions*, Cambridge, MA: Cambridge University Press.
Parsons, T. (1980) *Nonexistent Objects*, New Haven: Yale University Press.
Recanati, F. (1993) *Direct Reference: From Language to World*, Oxford: Blackwell.
Reimer, M. (2001) 'The Problem of Empty Names,' *Australasian Journal of Philosophy* 79: 491–506.
Russell, B. (1905) 'On Denoting,' *Mind* 14: 479–93.
Russell, B. (1911) 'Knowledge by Acquaintance and Knowledge by Description,' *Proceedings of the Aristotelian Society* 11: 108–28.
Russell, B. (1919) *Introduction to Mathematical Philosophy*, London: George Allen & Unwin.
Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.
Salmon, N. (1998) 'Nonexistence,' *Noûs* 32: 277–319.
Salmon, N. (2002) 'Mythical Objects,' in J.K.Campbell, M. O'Rourke, and D. Shier (eds.) *Meaning and Truth: Investigations in Philosophical Semantics*, New York: Seven Bridges Press.
Sawyer, S. (2002) 'Abstract Artifacts in Pretence,' *Philosophical Papers* 31: 183–98.
Sawyer, S. (2010) 'The Modified Predicate Theory of Proper Names,' in S. Sawyer (ed.) *New Waves in Philosophy of Language*, London: Palgrave Macmillan Press.
Soames, S. (1987) 'Direct Reference, Propositional Attitudes, and Semantic Content,' *Philosophical Topics* 15: 47–87.
Soames, S. (2002) *Beyond Rigidity*, Oxford: Oxford University Press.
Strawson, P.F. (1950) 'On Referring,' *Mind* 59: 320–44.
Taylor, K. (2000) 'Emptiness Without Compromise,' in A. Everett and T. Hofweber (eds.) *Empty Names, Fictional Names, and the Puzzles of Nonexistence*, Stanford: CSLI Press.
Thomasson, A. (1998) *Fiction and Metaphysics*, Cambridge: Cambridge University Press.

1.14

RELEVANCE THEORY

Robyn Carston

1 Introduction

Relevance theory (RT) is best known for its account of verbal communication and comprehension, but it also sets out a general picture of the principles driving the human cognitive system as a whole and this plays a crucial role in underpinning the particular claims made about communication and the pragmatic theory that follows from them.

The various post-Gricean accounts of the principles and processes that mediate the gap between sentence meaning and speaker meaning can be divided broadly into three classes based on their orientation: linguistic, philosophical, and cognitive-scientific. Linguistically oriented theories tend to focus on those pragmatic processes which are the least context-sensitive and most code-like, reflecting default or general patterns of language use (Levinson 2000; Horn 1984, 2004). Philosophically oriented accounts tend to follow Grice closely in maintaining his system of conversational norms and providing rational reconstructions of the "conversational logic" that delivers speakers' implicated meaning (Neale 1993, chapter 3; Recanati 2001, 2004). Given its cognitive-scientific orientation, relevance theory pragmatics is concerned with the on-line processes of utterance interpretation and the nature of the mental system(s) responsible for them (Sperber and Wilson 1986/1995a; Wilson and Sperber 2004). So it is responsive to research in evolutionary psychology on the nature of human cognitive architecture, empirical work on children's communicative development and experimental measures of adults' on-line comprehension, investigations into the relation between pragmatic competence and theory of mind (the ability to attribute intentions and beliefs to others), and clinical studies of people with impaired communicative capacities. For a survey of the ways in which Relevance Theory engages with these issues, see Wilson and Sperber (2004), Wilson (2005).

Given the philosophical nature of this volume, I will focus less in this chapter on the cognitive theorizing and experimental work that has built up around relevance theory than on those issues which have brought it into direct contact with debates in the philosophy of language. These include the meaning and function of singular terms (names, indexicals, demonstratives) and definite descriptions, the apparent occasion-sensitivity of word meaning and the extent to which pragmatics may affect the truth-conditional content of an utterance. These are all issues that bear on the distinction between the meaning provided by the linguistic system and the meaning that arises through the pragmatics of human communicative interaction. Broadly speaking, philosophers of language fall into two camps: *semantic minimalists*, who maintain that natural language

sentences provide a propositional content that is essentially pragmatics-free, and *semantic contextualists*, who insist that it is only utterances (or speech acts) that express propositional contents and these are irremediably context-sensitive. Although RT is usually classified as a contextualist theory, it will be suggested that, on the basis of its cognitive underpinnings and its emphasis on minds in communication, it occupies a distinct position, which I call "pragmaticism." These philosophically oriented issues are taken up in section 3, but first, in section 2, I lay out the main tenets of the theory.

2 Relevance Theory—Principles and Processes

2.1 Relevance Theory and Cognition

According to the RT framework, human cognitive systems quite generally are geared towards achieving as many improvements to their representational contents and to their organization as possible, while insuring that the cost to their energy resources is kept as low as reasonably possible. At the center of the theory is a technically defined notion of *relevance*, where relevance is a potential property of any input to any perceptual or cognitive process. An input may deliver a variety of different types of *cognitive effects* to the system; it may, for instance, combine inferentially with existing assumptions to yield new conclusions (known as *contextual implications*), it may provide evidence that strengthens existing beliefs, it may contradict and eliminate already held information, or it may rearrange the way information is stored. Such effects may or may not be beneficial to an individual, that is, they may increase or decrease the accuracy of the cognitive system's information about the world and may make useful information easier or harder to access. An input is relevant to a cognitive system only if it benefits that system, that is, only if it has *positive* cognitive effects. The other crucial factor affecting the degree of relevance of an input (whether an external stimulus or an internal mental representation) is the *processing effort* it consumes: Deriving effects from any given input requires a mobilization of cognitive resources, including attention, memory, and various processing algorithms and heuristics. Thus, the relevance of any input is a trade-off between the positive cognitive effects it yields and the processing effort it requires: the greater the ratio of effects to effort the greater the relevance of the input. The basic claim of the framework is that human cognition is oriented towards maximizing relevance (known as the *Cognitive Principle of Relevance*). The evolutionary idea underlying this is that, as a result of constant selection pressure towards increasing cognitive efficiency, humans have evolved procedures to pick out potentially relevant inputs and to process them in the most cost-effective way (Sperber and Wilson 1995).

RT has developed in tandem with two important and closely connected ideas in cognitive science: (1) that the mind is modular, and (2) that many mental processes are performed by fast and frugal heuristics. Arguments from evolutionary psychology suggest that the mind is massively modular, in the sense that a great many distinct dedicated procedures and processes have evolved to solve specific cognitive problems (Cosmides and Tooby 1994; Sperber 2002). And it seems that, in solving a wide variety of everyday problems, we employ rather simple, albeit ecologically rational, heuristics rather than foolproof algorithms or explicit reasoning processes (Gigerenzer et al. 1999). These mechanisms making up the mind's "adaptive toolbox" are "fast and frugal" in that they carry out limited computations and consult just a small salient subset of all the available information rather than working through myriad possibilities and comparing candidate

solutions. A process of this sort can only achieve a high degree of accuracy in a particular domain to whose regularities it is specifically tailored or adapted.

The relevance-theoretic approach to communication situates pragmatics within this sort of cognitive framework, that is, one which consists of largely domain-specific capacities, each with the function of solving a specific pressing problem in human mental life and employing quick, relatively cheap computations to do so (Sperber and Wilson 2002; Allott 2008). The human pragmatic capacity is such a dedicated system: Its specific domain is ostensive stimuli (verbal utterances and other acts of ostensive communication) and the comprehension procedure it employs is a fast and frugal heuristic. This idea is elaborated in the next section.

2.2 Relevance Theory and Linguistic Communication

The starting point for a pragmatic theory is the question of how hearers are able to bridge the gap between the linguistic meaning encoded in an utterance and the speaker's intended meaning. The most obvious manifestations of this gap are nonliteral uses of language such as metaphor and irony, and cases where, as well as the proposition she explicitly expresses, the speaker communicates an additional proposition known as a conversational implicature, exemplified by speaker Y's utterance in (1).

(1) X: We need your written report now.
Y: I've been very busy recently.
Implicating: I haven't written the report yet.

There is also a range of pragmatic tasks involved in determining the proposition explicitly expressed, including disambiguation, assignment of referents to indexicals, and filling in missing constituents, as in (2), and various other enrichments or adjustments of encoded content, as indicated in the examples in (3) (where the particular proposition expressed given here is, of course, just one of indefinitely many possibilities):

(2) He has taken enough from her.
Expressing: Jim has endured enough abusive treatment from Mary.
(3) a. I've eaten.
Expressing: I've eaten dinner tonight.
b. Your knee will take time to heal.
Expressing: Your knee will take a substantial amount of time to heal.
c. The water is boiling.
Expressing: The water is very hot [not necessarily strictly at boiling point].

While the proposition expressed by (3a) contains constituents of content that don't appear in the sentence uttered, (3b) involves a narrowing down of "take time" and (3c) a loosening of the concept encoded by "boiling."

How, then, is an addressee able to infer the intended meaning from the encoded linguistic meaning which is just a schematic guide or set of clues? According to RT, the answer lies with a special property of overtly communicative acts: which is that they raise certain expectations of relevance in their addressees—that is, expectations about the cognitive effects they will yield and the mental effort they will cost. Quite

generally, an utterance comes with a presumption of its own *optimal relevance*—that is, there is an implicit guarantee that the utterance is the most relevant one the speaker could have produced, given her competence and her own current goals, and that it is at least relevant enough to be worth processing. This is known as the *Communicative Principle of Relevance* and it follows from the Cognitive Principle of Relevance in conjunction with the overtness of the intention that accompanies an utterance: The speaker openly requests effort (attention) from her addressee who is thereby entitled to expect a certain quality of information requiring no gratuitous expenditure of effort. That utterances carry this presumption licenses a particular comprehension procedure, which, in successful communication, reduces the number of possible interpretations to one:

Relevance-theoretic comprehension procedure:

a. Follow a path of least effort in computing cognitive effects: Test interpretive hypotheses (disambiguations, reference resolutions, lexical adjustments, implicatures, etc.) in order of accessibility.
b. Stop when your expectations of relevance are satisfied.

This procedure (a fast and frugal heuristic) is automatically applied in the on-line processing of verbal utterances: Taking the schematic decoded linguistic meaning as input, processes of pragmatic completion and enrichment at the explicit level occur in parallel with the derivation of the implications of the utterance. Central to the working of the procedure is a subprocess of *mutual adjustment* of explicit content and contextual implications, a process guided and constrained by expectations of relevance. Here is a brief example involving the adjustment of explicit content in response to expected implications and where the outcome is a narrowing down of a lexically encoded meaning:

(4) Bill: I'm doing the 10km circuit run this afternoon. Would you like to come?
Sue: No thanks, I'm *resting* today.

The verb "rest" encodes a rather general concept, REST, which covers any degree of inactivity (physical or mental), from sleeping, to staying awake but not moving much, to performing a range of not very strenuous tasks (with many more possibilities in between). Suppose now that Sue is quite an athletic person, who exercises regularly, then her use of "rest" here is plausibly understood as expressing a concept REST*, which entails a much lower degree of physical activity than she undertakes on her training days but is still quite compatible with her pottering about the garden or walking to the shops. A hearer using the relevance-theoretic comprehension heuristic would narrow the encoded concept REST just as far as is required to satisfy his expectation of relevance (e.g. by explaining why Sue is refusing his invitation), and no further. This particular narrowing would cost Bill little effort, given his knowledge of Sue's exercise habits, and provide him with a range of contextual implications (e.g. she won't come with me today because she is RESTING*, she may come another day when she isn't RESTING*, etc.). In different circumstances—for instance, in response to the question, "Would you like to walk to the corner shop with me?"—REST would have to be narrowed much further, resulting in a distinct concept REST**.

An interesting RT claim in this context is that metaphoric and hyperbolic uses of words involve a kind of concept broadening ("loose use"), so fall within this general process of lexical meaning adjustment, which contributes to explicitly communicated content. For instance, an utterance of the sentence in (5) could be taken literally, or as an approximation (if, say, the run referred to was a little less than 26 miles), or as hyperbolic (if it was obviously much less than the length of a marathon), or as metaphorical for a long, arduous, exhausting experience, whether physical or mental. The idea is that there is a continuum of degrees (and kinds) of concept broadening, with metaphorical use being the most radical case.

(5) It was a *marathon*.

For detailed exemplification of the RT-based account of lexical adjustment, resulting in concept broadening, or narrowing, or a combination of the two, see Carston 2002, Wilson and Sperber 2002, and Wilson and Carston 2007.

It is a basic assumption of RT pragmatics that the meaning encoded in the linguistic expression type uttered inevitably underdetermines the content that speakers communicate, not only their implicatures but also the propositional content they communicate explicitly (the explicature of their utterance). It is claimed that this is not simply a matter of effort-saving convenience on the part of speakers who could employ fully explicit (eternal) sentences if they so chose, but rather it is an intrinsic property of public language systems and no amount of effort to be more explicit by employing ever longer and more complex sentences will achieve full explicitness (Carston 2002 chapter 1, 2009). Since an act of ostensive communication, verbal or nonverbal, preempts the addressee's attention and triggers inferential processes geared toward finding an optimally relevant interpretation of the act, much of the speaker's intended meaning can be communicated without being encoded. Assuming that language evolved in the crucible of an inferential pragmatic system already employed in preverbal communication, the function of linguistically encoded meaning is to channel the inferential mechanism in particular directions and to provide just that content that cannot be recovered by inference alone (Sperber 2000; Origgi and Sperber 2000).

In the next section, implications of this view of the language/pragmatics relation are considered for some central semantic issues within current philosophy of language.

3 Relevance Theory and Issues in Philosophy of Language

3.1 Reference and Definite Descriptions

Virtually everyone agrees on the inherently context-sensitive nature of indexical and demonstrative expressions and, therefore, on the fact that understanding utterances containing them requires a process of pragmatic saturation. However, there is much less agreement on the role of pragmatics in grasping the content of definite descriptions and proper names in use. According to recent work in Relevance Theory, although they are not linguistically marked as context-sensitive and so as requiring pragmatic saturation, their linguistically encoded meaning does not determine their reference and pragmatics plays an essential role (see, in particular, Powell 2001, 2003, 2010).

As is well known, the key question on definite descriptions over the last 50 years has been how to analyze what Donnellan (1966) calls the 'referential-attributive

distinction', that is, the fact that a definite description, "the F," may be used either to talk about a particular identifiable individual or to talk about *whatever/whoever* happens to be uniquely F. As regards proper names, there are broadly three positions: those who consider names to be the natural language equivalent of logical individual constants, those who take them to be descriptive, that is, to contribute properties to truth conditions, and those who see them as closely related to indexicals.

How best to treat the referential-attributive distinction has proved a contentious issue within the philosophy of language. On the one hand, the truth conditions of an utterance of a definite description sentence appear to alter according to whether the description is used referentially or attributively. On the other hand, there seem to be good reasons to believe that definite descriptions are not linguistically ambiguous: Although the description 'the mayor of this town' may be used either referentially or attributively, it does not seem to be ambiguous in the way that words like "bank" or "match" are. But these two observations are hard to reconcile on standard philosophical assumptions: if (leaving aside indexical expressions) you identify the meaning of an expression with the contribution that expression makes to truth conditions, then it follows directly that an expression which is capable of making two different kinds of contribution to truth conditions has two different meanings and so is ambiguous.

However, Relevance Theory offers a natural way to reconcile these data since it draws a clear distinction between, on the one hand, the linguistically encoded meaning of a particular expression and, on the other hand, the contribution that expression makes to truth-conditional content on an occasion of use. Given this distinction, the fact that a particular expression may make two (or more) distinct kinds of contribution to truth-conditional content is not evidence for its ambiguity at the level of linguistically encoded meaning. According to the relevance-theoretic account, definite descriptions are linguistically univocal, as Donnellan himself maintained, but are truth-conditionally ambiguous (Bezuidenhout 1997; Powell 2001). The gap between the encoded meaning of a definite description and what that description contributes to the propositional content of a particular utterance is bridged by relevance-guided pragmatic inference.

There has been less work from a relevance-theoretic perspective on proper names than on definite descriptions, but Powell (1998, 2003, 2010) has addressed questions about proper names as part of a general relevance-theoretic analysis of the semantics and pragmatics of singular expressions. According to his analysis, all such expressions are profoundly context-sensitive: whether they make referential or descriptive contributions to truth conditions is not a matter of the encoded meanings of these expressions, but is rather a matter of broad context and pragmatic principles. Powell analyses the encoded meanings of singular expressions (including here definite descriptions) not in terms of their contribution to truth conditions, but rather in terms of their contribution to a hearer's mental representations. All these expressions, on this view, are marked as *individual concept communicators* by virtue of their linguistically encoded meaning. That is to say, their encoded meaning indicates that what they contribute to explicature is a concept which is taken to be satisfied by a unique individual, but they are neutral with regard to whether this concept should be *de re* (referential) or descriptive (attributive)—this has to be determined pragmatically. Which constraint a particular singular expression places on the concepts which may serve as its interpretation varies according to the type of singular expression. In the case of a proper name "N," the constraint on interpretation is simply that the individual concept should be of a bearer of "N." Which concept that is on a particular occasion of use will be determined by pragmatic

inference, constrained by the criterion of optimal relevance. A definite description "the F," on the other hand, encodes a more complex condition: It constrains interpretation to an individual concept of a unique F in a salient context. Again, which is the salient context and which the intended individual concept (and whether it is referential or descriptive) on any given occasion is determined pragmatically.

On this account of singular terms, the encoded linguistic meaning provides a set of indications concerning the conceptual content to be recovered, but, on all occasions of use, some pragmatic inference is required for the full determination of that content. For a detailed account of how, following the relevance-based comprehension procedure, either the referential (*de re*) or the attributive (descriptive) interpretation is recovered, see Powell (2001).

3.2 'Free' Pragmatic Processes and Unarticulated Constituents

Ascertaining the proposition that a speaker explicitly communicates (the explicature) is a matter for pragmatics. A controversial claim in this context is that there are "free" pragmatic processes that can affect this level of content, where "free" is understood as not required or directed by any element of the linguistic expression used; see Recanati (1993: 243) for a clear distinction between linguistically controlled "saturation" processes and linguistically free pragmatic enrichment processes.

There are, arguably, two kinds of "free" enrichment processes. The first of these is the modulation or adjustment of lexically encoded meanings, as exemplified above in (4) and (5). The second, and more contentious, is the pragmatic recovery of components of content which are not linguistically indicated in any way (and so are known as "unarticulated constituents" of utterance content). The following are some cases for which relevance-theorists have claimed that the bracketed constituent, a component of the explicature of a particular utterance of the sentence, is not represented in the linguistic form:

(6) a. It is raining. {in Oslo}
b. I have eaten. {supper}
c. Jill reported Jack for misconduct and he was fired. {as a result}

It seems that, in comprehending an utterance of (6a), we would very often, if not always, take there to be a specific place at which the raining is occurring; this would often be the location of the speaker of the utterance, but need not be. Similarly, for many utterances of (6b), an object of eating is recovered and, for (6c), a cause-consequence relation is taken to hold between the two events described in the conjuncts. According to the view that these are linguistically unarticulated constituents of the propositions expressed by speakers of these utterances, their recovery by addressees is entirely pragmatically motivated; that is, they are inferred as part of the process of finding an optimally relevant interpretation.

This kind of "strong" pragmatic effect on the proposition explicitly expressed by an utterance is resisted by semanticists such as Stanley (2000, 2005), King and Stanley (2005) and Martí (2006). They see this propositional content as the semantics (the truth-conditional content) of the sentence, relative to a context of utterance, and insist that, as such, all its constituents must be provided for in the linguistic logical form (LF) of the sentence. Pragmatics can only have "weak" effects at this level of content; that is,

its role is just to supply context-specific values to indexical elements in the logical form (the saturation process). As well as overt, phonologically realized, indexical elements such as pronouns and demonstratives, a sentence may contain various covert indexicals which, although not phonologically manifest, constitute elements of syntactic structure. So, on this view, the sentences in (6a) and (6b) contain such elements; for instance, (6a) includes an unpronounced adjunct of location, <at L>, which indicates that a value is to be pragmatically supplied in just the same way as it is to an occurrence of the overt demonstratives "here" or "there." Stanley (2000) finds support for the presence of this covert location indexical from the fact that, like an overt pronoun, it can enter into binding relations, as in "Every time John lights a cigarette, it rains," on the interpretation where the location of each event of raining is bound to the location of each event of John's lighting a cigarette. In the case of (6c), on the other hand, no plausible case can be made for a covert indexical and it is accepted that the cause-consequence relation is a strong pragmatic effect (the result of a "free" pragmatic process). On that basis, it is treated as an instance of conversational implicature, hence as falling outside the truth-conditional content of the utterance (King and Stanley 2005), contrary to widespread intuitions.

In a critical response to this "indexicalist" semantic account, relevance-theorists have pointed out that there is no principled limit to the number of covert elements that such a theory would have to posit (Wilson and Sperber 2002). An utterance of (6b), for instance, might express a proposition with a range of constituents corresponding to what is eaten, the time, place, manner of eating, and so on. On Stanley's analysis, each of these would have to correspond to a variable or indexical in the linguistic logical form, a theoretical prediction which Wilson and Sperber take to be a *reductio* of his position. Furthermore, as Carston (2002) points out, although all of these hidden indexical elements would have to be present at LF, there would be many occasions of utterance of these sentences on which some of these elements would receive no value. Consider an utterance of (6b) "I've eaten" in response to the question "Would you like some dinner?" While what was eaten and the time of eating might well be relevant (that the speaker has eaten a full meal and the eating took place in the recent past), the place and manner of eating would surely not be and would not receive a contextual value. However, this seems to be quite different from the way that overt indexicals work:

(7) She put the book on the table.

If the addressee of an utterance of the sentence in (7) is unable to assign a value (a referent) to "she," he has not fully understood the speaker's intended content and something has gone wrong with the communicative exchange. So it seems that Stanley's hidden indexicals are not, after all, simply covert counterparts of the overt indexicals that we are familiar with but entities of some other, unknown sort. The suspicion, from the pragmaticist side, is that they are an artefact of the semantic theorizing and don't actually exist.

Some within the semanticist camp have taken this objection seriously and looked for a remedy consonant with the position that any pragmatic processes affecting this level of propositional content are linguistically controlled. Thus, Martí (2006) has argued that the problem can be overcome by taking the covert indexical elements to be optional, so that, if and when they do occur, they must be pragmatically saturated, as is the case for their overt counterparts. Thus "it is raining" has two underlying logical forms, one

with a covert location indexical, one without. Clearly, this move entails a considerable increase in linguistic structural ambiguity. For instance, the surface sentence form in (6b) "I have eaten" would have a variety of underlying logical forms, each with an array of covert indexicals, differing in number and type (including one with none), indicating possible contextual completions. In the case of a sentence that has four possible covert indexicals for different constituents, the optionality position results in 16 logical forms to cover the range of cases. Thus, the shift from obligatory to optional linguistic structure comes at a high computational cost. While the semanticist acknowledges this point, she finds it preferable to positing a "mysterious and computationally intractable" process of free pragmatic enrichment (Martí 2006: 151).

However, according to Carston (2010), a counter-intuitive proliferation of linguistic ambiguity is not the only problematic outcome of making covert indexicals optional. When we spell out the implications of the optionality move for the on-line processes of comprehension, it seems that the logical forms with their varying numbers of covert indexicals are redundant. The semanticists' requirement is that indexical saturation, rather than a free pragmatic process, is responsible for any unpronounced constituents that occur in the explicature of utterances of sentences like (6a) and (6b). So a logically prior task, on any given occasion of utterance, is disambiguation of the surface phonological form, so that one among the range of logical forms with their varying configurations of indexicals is selected as the intended one. But, according to Carston, this disambiguation process relies on the recovery of relevant contextual information (e.g. the specific location of raining or the object of eating) which is identical to the constituents of content which would be supplied directly by a free pragmatic process as part of the inferential construction of an optimally relevant interpretation. If this is right, a wholly pragmatically driven process of inferring a constituent of content preempts any process of indexical saturation and renders superfluous the logical forms containing covert indexicals.

For further relevance-theoretic arguments against hidden indexicals (whether obligatory or optional) and a response to allegations that free pragmatic enrichment is insufficiently constrained, see Hall (2008a, 2008b, 2009).

3.3 Relevance Theory and the Minimalism/Contextualism Debate

The linguistic meaning of the sentence a speaker utters is decoded and processed bit by bit in on-line comprehension, but thinking hypothetically of it as a whole, what it amounts to, on the RT view, is a template or schema for constructing the proposition the speaker explicitly communicates. It is seldom, if ever, fully propositional (truth-conditional) and has no reason to be so, since the propositional content which is communicatively intended by the speaker is inevitably recovered by pragmatic inference. A number of philosophers of language concur with this view of what is delivered by "semantic" interpretation (characterized by its formal, algorithmic, context-free character). For instance, Recanati (2001) thinks of linguistic meaning as providing semantic schemata or propositional functions, and Bach (1994, 2006) talks of "propositional radicals." Both of them maintain that pragmatic processes (including "free" enrichment) make an essential contribution to retrieving the communicatively intended propositional content. Thus, on all of these accounts, there is a fairly obvious sense in which the encoded linguistic meaning is "minimal": it underdetermines the proposition explicitly communicated and is standardly less than fully propositional itself.

However, those philosophers of language who advocate a *minimal semantics* for natural language sentences see themselves as being in direct opposition to relevance-theorists and to "contextualist" philosophers such as Recanati and Bach. The most prominent current minimalists are Borg (2004, 2007) and Cappelen and Lepore (2005) and their primary concern appears to be to establish that uttered tokens of natural language sentences "semantically express" a proposition (a truth-conditional content) which is essentially context-invariant and pragmatics-free. While Cappelen and Lepore allow for a limited degree of context-relativity, restricted to the small class of overt indexicals and demonstratives, Borg (2004) argues that even these receive a semantic value without recourse to pragmatics or broad context. What is meant here by "minimalism," then, is *propositional* minimalism. It is a stance that follows from the longstanding assumption that what a semantics for any representational system must deliver is statements of truth-conditional content. However, it is difficult to see why we should accept this assumption, or why the minimalists are so attached it, since they agree with relevance-theorists that this "semantically expressed" proposition is seldom, if ever, the proposition explicitly communicated by an utterance. They accept that communicated propositions (both explicatures and implicatures) are pragmatic entities, derived by nondemonstrative inferential processes geared toward recovering an interpretation that meets expectations of rational communicative behavior, whether Gricean maxims or the presumption of optimal relevance (see, in particular, Borg 2004: 110). It seems clear, then, that while what is communicated (by uttering a declarative sentence) must be propositional (must embody statements or claims about the world, which an addressee may or may not take on as part of his own belief system), what is merely semantically expressed or encoded need not be. (For more detailed RT-oriented discussion of Cappelen and Lepore's "insensitive" semantics and of Borg's semantic minimalism, see Carston 2008a and 2008b, respectively.)

The two central tenets of contextualist semantics are: (a) that it is not sentences but utterances (or speech acts) that have truth-conditional content, and (b) that virtually every word in the language is context-sensitive. See, for instance, Searle (1978, 1980), Recanati (1993, 2004) and Travis (1985, 1997). Unsurprisingly, then, Relevance Theory is usually placed squarely within the contextualist camp—by minimalists and contextualists alike.

However, it's not clear that this is the best way to characterize Relevance Theory nor that this distinction between *semantic* theories is central to what RT is about. It is first and foremost a theory of communication and interpretation, and its advocacy of free pragmatic processes (meaning modulation and unarticulated constituents) is entirely motivated by the aim of providing an account of how it is that hearers can recover speakers' intended contents, given that these diverge in a range of ways from the meaning encoded in the linguistic expressions employed. What contextualists take to be the semantics (the truth-conditional content) of an utterance is, for relevance theorists, the speaker's explicature—that is, the proposition she explicitly communicates. Since thoughts have truth-conditional content, it follows that any *communicated thought* has truth conditions, whether it is explicitly or implicitly communicated (an explicature or an implicature). So, on this account, which posits a nonpropositional linguistic semantics and, via pragmatics, a set of communicated propositions, there appears to be no reason to think of the explicature of an utterance as its semantics.

RT agrees with contextualist philosophers that the meaning a use of a word contributes to an explicature varies from occasion to occasion. Virtually any expression can be used by a speaker and understood by a hearer to express a meaning that is different from its linguistic expression-type meaning. For instance, a speaker can use the word "butterfly" to communicate a concept whose denotation includes human beings who have certain properties, or the word "bachelor" to communicate a concept whose denotation includes some married men and excludes some unmarried men. But this is a rather different phenomenon, involving a different property of the words concerned, from the context-sensitivity of indexicals and demonstratives, which arguably do not encode a concept to start with but rather a variable with certain indications about the kind of value that variable should receive. Cappelen and Lepore (2005) have some quite effective tests for distinguishing this latter class of linguistic expressions from the rest and it comes as no surprise that there are such discriminatory tests: There are strong pretheoretic intuitions that indexicals are special and quite different from words like "butterfly" or "bachelor."

Rather than taking the contextualist stance that virtually all words are context-sensitive, RT makes a distinction between inherent context-sensitivity, on the one hand, (and agrees that it is confined to pretty much the cases that Cappelen and Lepore cite and which pass their various tests), and what could be called "*pragmatic susceptibility*," on the other hand. What this means is that virtually every (open class) word can be used and understood to express a meaning that departs in certain ways from the meaning that it encodes (its expression-type meaning) and this is because of the human pragmatic interpretive ability (which includes an acute sensitivity to relevant contextual factors). Linguistic expressions are tools with certain inherent properties (phonological, syntactic, and semantic) that we, as normally functioning adult humans, can employ very flexibly for our communicative purposes by virtue of certain characteristics of our psychological makeup (specifically, our "theory of mind" capacities, and, in particular, our attunement to each other's communicative intentions and our expectations of each other as rational speakers and hearers). Thus, while there is a limited degree of context-sensitivity built into linguistic systems, pragmatic susceptibility is a pervasive feature of language as employed by us in ostensive communication. This is one respect in which relevance theory is better described as being "radically pragmaticist" rather than "radically contextualist": It is us, the users of language, that are sensitive to context, and, as rational communicating/interpreting agents, we are able, by exploiting this sensitivity in each other, to make linguistic expressions do a lot more than simply express their standing linguistic meaning.

With regard to their views on the role of context, there is another way in which contextualist and pragmaticist orientations differ. In his pioneering work on conversational logic, Grice set out an inferential schema for deriving conversational implicatures which employed a large component of theory-of-mind type reasoning. However, he did not carry this over to the pragmatic processes required for the full identification of what a speaker has said (explicitly communicated) but spoke here of "context as a criterion." It seems that he thought of disambiguation and indexical reference assignment as a matter of contextual best fit, rather than as involving conversational maxims or processes of reasoning geared to the recovery of what the speaker intended (Grice 1989: 25, 222). In this regard, certain current contextualists, despite being strong advocates of free pragmatic processes of enrichment and modulation in recovering what a speaker has said, remain essentially Gricean. For instance, while Recanati (2004) construes the

"secondary" pragmatic processes of conversational implicature derivation as maxim-guided, reflective reasoning, which deploys premises concerning speakers' mental states (beliefs and intentions), he takes a different stance on the "primary" pragmatic processes that contribute to the recovery of explicatures. He sees these as a function of an automatic, dumb (noninferential) cognitive mechanism responsive to differential degrees of activation of candidate interpretations, such that the most highly activated one wins out. It is context (both linguistic and extra-linguistic) that does the work here and contextual coherence that provides the criterion of correctness. As he puts it, "the interpretation which eventually emerges . . . results from a blind, mechanical process, involving no reflection on the interpreter's part. The dynamics of accessibility does everything and no 'inference' is required. In particular, there is no need to consider the speaker's beliefs and intentions" (Recanati 2004: 32).

According to relevance theory, on the other hand, the whole utterance interpretation process is a matter of (nondemonstrative) inference, and taking account of the speaker's competence (including her epistemic states) and preferences (her desires, intentions, interests) may be required for carrying out any of the pragmatic tasks involved (including lexical concept adjustments, disambiguation, fixing of indexical reference). Along with the propositions communicated (explicatures, implicatures), aspects of the context for the interpretation fall under the speaker's communicative intention and the hearer selects it (in the form of a set of conceptual representations) as part of his search for an interpretation that satisfies his expectations of relevance.

Summing up, the pragmaticism of relevance theory has the following characteristics which distinguish it from contextualism: (a) it is primarily a theory of communication and interpretation rather than a semantic theory and, although explicatures have truth-conditional content (as do implicatures, thoughts, and propositional entities quite generally), there is no useful sense in which they are the semantic content of anything (a sentence, a sentence token in a context, an utterance); (b) while only a few words in the language are inherently context-sensitive, the vast majority of words are susceptible to the pragmatics of the speaker–hearer interaction such that they can be used to communicate an indefinite range of different concepts; (c) it is not context acting on language that is somehow doing the work of determining explicature content, but, just as for implicatures, it is the exercise of specific human mind-reading capacities dedicated to achieving the feat of ostensive communication and comprehension.

Related Topics

1.2 Semantics and Pragmatics
1.6 Pragmatic Enrichment
1.7 Meaning and Communication
1.10 Context-Sensitivity
2.1 Reference
3.1 Names
3.7 Descriptions
4.3 The Role of Psychology
7.6 Ordinary Language Philosophy.

References

Allott, N. (2008) Pragmatics and Rationality. PhD thesis, University of London.

Bach, K. (1994) "Conversational Impliciture," *Mind and Language* 9: 124–162.

Bach, K. (2006) "The Excluded Middle: Semantic Minimalism without Minimal Propositions," *Philosophy and Phenomenological Research* 73 (2): 435–442.

Bezuidenhout, A. (1997) "Pragmatics, Semantic Underdetermination and the Referential/Attributive Distinction," *Mind* 106: 375–409.

Borg, E. (2004) *Minimal Semantics*. Oxford: Clarendon Press.

Borg, E. (2007) "Minimalism Versus Contextualism in Semantics," in G. Preyer and G. Peter (eds.) *Context-Sensitivity and Semantic Minimalism*. Oxford: Oxford University Press, 339–359.

Cappelen, H. and Lepore, E. (2005) *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.

Carston, R. (2002) *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.

Carston, R. (2008a) "Linguistic Communication and the Semantics/Pragmatics Distinction," *Synthèse* 165 (3): 321–345. (Special Issue on the Semantics/Pragmatics Distinction.)

Carston, R. (2008b) "A Review of E. Borg, 2004. *Minimal Semantics*. Oxford: Clarendon Press," *Mind and Language* 23: 359–67.

Carston, R. (2009) "The Explicit/Implicit Distinction in Pragmatics and the Limits of Explicit Communication," *International Review of Pragmatics* 1 (1): 35–62.

Carston, R. (2010) "Explicit Communication and 'Free' Pragmatic Enrichment," in B. Soria and E. Romero (eds.) *Explicit Communication: Robyn Carston's Pragmatics*. Basingstoke: Palgrave Macmillan, 217–287.

Cosmides, L. and Tooby, J. (1994) "Origins of Domain-Specificity: the Evolution of Functional Organization," in L. Hirschfeld and S. Gelman (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press, 85–116.

Donnellan, K. (1966) "Reference and Definite Descriptions," *Philosophical Review* 75: 281–304.

Gigerenzer, G., Todd, P.M., and the ABC Research Group. (1999) *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

Hall, A. (2008a) "Free Enrichment or Hidden Indexicals?" *Mind and Language* 23(4): 426–456.

Hall, A. (2008b) Free Pragmatic Processes and Explicit Utterance Content. PhD thesis, University of London.

Hall, A. (2009) "'Free' Enrichment and the Nature of Pragmatic Constraints," *UCL Working Papers in Linguistics* 21: 93–123.

Horn, L. (1984) "Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature," in D. Schiffrin (ed.) *Meaning, Form and Use in Context (GURT '84)*.Washington, DC: Georgetown University Press, 11–42.

Horn, L. (2004) "Implicature," in L. Horn and G. Ward (eds.) *Handbook of Pragmatics*, Oxford: Blackwell, 3–28.

King, J. and Stanley, J. (2005) "Semantics, Pragmatics, and the Role of Semantic Content," in Z. Szabo (ed.) *Semantics vs. Pragmatics*, Oxford: Oxford University Press, 111–164.

Levinson, S. (2000) *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Martí, L. (2006) "Unarticulated Constituents Revisited," *Linguistics and Philosophy* 29: 135–166.

Neale, S. (1993) *Descriptions*. Cambridge, MA: MIT Press.

Origgi, G. and Sperber, D. (2000) "Evolution, Communication and the Proper Function of Language," in P. Carruthers and A. Chamberlain (eds.) *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*. Cambridge: Cambridge University Press, 140–169.

Powell, G. (1998) "The Deferred Interpretation of Indexicals and Proper Names," *UCL Working Papers in Linguistics* 10: 143–172.

Powell, G. (2001) "The Referential-Attributive Distinction—a Cognitive Account," *Pragmatics and Cognition* 9 (1): 69–98.

Powell, G. (2003) Language, Thought and Reference. PhD thesis, University of London.

Powell, G. (2010) *Language, Thought and Reference*, Basingstoke: Palgrave Macmillan. (Substantially revised version of Powell 2003.)

Recanati, F. (1993) *Direct Reference: From Language to Thought*. Oxford: Blackwell.

Recanati, F. (2001) "What is Said," *Synthèse* 128: 75–91.

Recanati, F. (2004) *Literal Meaning*.Cambridge: Cambridge University Press.
Searle, J. (1978) "Literal Meaning," *Erkenntnis* 13: 207–224.
Searle, J. (1980) "The Background of Meaning," in J. Searle, F. Keifer ,and M. Bierwisch (eds). *Speech Act Theory and Pragmatics*. Dordrecht: Reidel, 221–32.
Sperber, D. (2000) "Metarepresentations in an Evolutionary Perspective," in D. Sperber (ed.) *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press, 117–137.
Sperber, D. (2002) "In Defense of Massive Modularity," in E. Dupoux (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge MA: MIT Press, 47–57.
Sperber, D. and Wilson, D. (1986/95a) *Relevance: Communication and Cognition*. Oxford: Blackwell.
Sperber, D. and Wilson, D. (1995) "Postface," in *Relevance: Communication and Cognition*, Oxford: Blackwell. Second edition 1995a, 255–279.
Sperber, D. and Wilson, D. (2002) "Pragmatics, Modularity and Mind-Reading," *Mind and Language* 17: 3–23. (Special Issue on Pragmatics and Cognitive Science.)
Stanley, J. (2000) "Context and Logical Form," *Linguistics and Philosophy* 23: 391–434.
Stanley, J. (2005) "Semantics in Context," in G. Preyer and G. Peter (eds.) *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Oxford University Press, 221–253.
Travis, C. (1985) "On What is Strictly Speaking True," *The Canadian Journal of Philosophy* 15 (2): 187–229. Reprinted in C. Travis 2008, 19–64.
Travis, C. (1997) "Pragmatics," in B. Hale and C. Wright (eds.) *A Companion to the Philosophy of Language*. Oxford: Blackwell, 87–107. Reprinted in C. Travis 2008, 109–129.
Travis, C. (2008) *Occasion-Sensitivity: Selected Essays*. Oxford: Oxford University Press.
Wilson, D. (2005) "New Directions for Research on Pragmatics and Modularity," *Lingua* 115: 1129–1146.
Wilson, D. and Carston, R. (2007) "A Unitary Approach to Lexical Pragmatics: Relevance, Inference and Ad Hoc Concepts," in N. Burton-Roberts (ed.) *Advances in Pragmatics*, Basingstoke: Palgrave Macmillan, 230–260.
Wilson, D. and Sperber, D. (2002) "Truthfulness and Relevance," *Mind* 111: 583–632.
Wilson, D. and Sperber, D. (2004) "Relevance Theory," in L. Horn and G. Ward (eds.) *The Handbook of Pragmatics*, Oxford: Blackwell, 607–632.

1.15

TRUTH AND REFERENCE IN FICTION

Stavroula Glezakos

Introduction

Fiction is often characterized by way of a contrast with truth, as, for example, in the familiar couplet "Truth is always strange/Stranger than fiction" (Byron 1824). And yet, there are those who maintain that we are likely to "learn more about human life and human personality from novels than from scientific psychology" (Chomsky 1988: 159); on their view, some truth is best encountered via fiction. The scrupulous novelist points out that her work depicts no actual person, either living or dead; nonetheless, we use names from fiction in ways that suggest that we take these names to refer. Philosophers who investigate fiction aim to reconcile such apparently incompatible phenomena, and, in general, to account for the myriad ways that we talk, think, and feel about fiction.

Fictional Characters and Fictional Names

Of great interest to philosophers has been the way that certain statements appear to be true in virtue of fiction. F. Scott Fitzgerald's novel *The Great Gatsby* (1925) contains descriptions of protagonist Jay Gatsby's lavish West Egg mansion; it thus seems appropriate to classify the following as true and false, respectively:

(1) Jay Gatsby lives in a lavish mansion in West Egg.
(2) Jay Gatsby lives in a polyurethane yurt near Big Sur.

The most straightforward account of truth for a sentence that contains a proper name and a predicate holds that the sentence is true just in case the referent of the name stands in the appropriate relation to the property expressed by the predicate. Thus, to maintain that (1) is true, one must either identify a referent for the name "Jay Gatsby," or provide an alternative account of the sentence's truth.

Many philosophers do, in fact, accept that the name "Jay Gatsby" refers. They disagree, however, about what it refers to, with some maintaining that the referent of "Jay Gatsby" is not a man, others, that it is a man that is not actual, and yet others, that it is something that does not exist.

Fictional Names Refer to Abstract Objects

According to abstract object theorists, fictional names refer to fictional characters, and critical assessment of a fictional work involves reference to such characters (Searle 1975, van Inwagen 1977, Salmon 1998, and Thomasson 1999, 2003, 2009). Consider, for example, the following sentence, which one might utter during a discussion of the literary merits of *The Great Gatsby*:

(3) Daisy is a less well-developed character than Gatsby.

These theorists maintain that accepting (3) commits one to the truth of (4):

(4) Some fictional character exists.

This is because from (3) we can infer that there exists a fictional character that is less well developed than Gatsby, and thus, that there exists a fictional character (van Inwagen 1977). Such characters—or "creatures of fiction"—are abstract entities, possessing (or lacking) the sorts of properties that are at issue within critical discourse, such as being well developed, being sympathetic, or being modeled on a particular person (van Inwagen 1977: 302). In uttering (3), one refers to two such characters, and it is their relative degree of development that determines whether one's utterance is true or false.

Matters become more complicated when we consider sentences like (1) and (2). If the referents of fictional names are abstract objects—that is, objects that are neither physical nor causally efficacious—then descriptions within novels cannot be specifying properties of those objects. And, indeed, abstract object theorists direct us to give a revised reading to any predicative sentence involving a fictional name and predicate assigned to a character within a novel. An assertive utterance of (1), for example, should be understood as the claim that *in the novel The Great Gatsby, the property of living in a lavish mansion in West Egg is ascribed to the character of Gatsby*—which these theorists take to be true.

Some abstract object theorists hold that fictional characters come into existence along with the particular work of fiction in which they figure (Searle 1975, Salmon 1998, and Thomasson 1999). On their view, creation of a work of fiction involves pretended reference to persons, places, and things; by means of such pretence, an author brings into being not only her novel, but also fictional persons, places, and things. Once created, a character can appear in a later work by someone other than its creator, and may be portrayed in ways that depart from the depictions offered in the original work, so long as the new author's intentions are appropriately directed at the character from that earlier story.

In contrast, Wolterstorff (1980) argues that the authorial task of character development is best understood as providing specifications for a *sort* of thing, rather than creation of a particular thing. On his account, names from fictional works refer to abstract universals, or "person-kinds," that exist prior to, and independently of, those works. One such kind consists of the conjunction of all and only the properties that are, in *The Great Gatsby*, assigned to Gatsby; a sentence like (1), if used to assert that *the Gatsby-kind is partly constituted by the property of living in a lavish mansion in West Egg*, comes out not just true but necessarily true. Understanding characters in this way allows us to suggest, of an actual person who happens to possess a large number of the features that

constitute the Gatsby-kind, that he strikes us as an instance of that kind (although, since no actual human could satisfy all and only the descriptions of Gatsby in the novel, Gatsby-hood could never actually be instantiated).

Perhaps unsurprisingly, the proposal that fictional characters are existing abstract entities has met with some objection. One concern is that it does not seem to provide for the truth of sentences like (5) and (6):

(5) Daisy has a larger wardrobe than Nick Carraway.
(6) Either Daisy has more freckles on her left leg than on her right, or she has more freckles on her right leg than on her left, or she has exactly the same number of freckles on both legs.

Since neither the size of Daisy's wardrobe relative to Nick's, nor the number of freckles on Daisy's legs, are discussed within *The Great Gatsby*, the sentences are false if given the reading *in The Great Gatsby*, . . . And, since abstracta do not have wardrobes or freckles, the sentences will also be false if they are read as direct predications of the fictional characters Nick Carraway and Daisy Buchanan. Thus, if we think that it is possible to make true claims about characters that go beyond the descriptions provided in the fiction, we may find ourselves unsatisfied by this account.

A deeper worry is that such fictional characters appear to be precariously balanced between inconsistency and superfluity. The characters that figure in the beliefs, intentions, and emotional responses of authors and readers (see Thomasson 1999; cf. Braun 2005) do not seem to be abstracta. Take, for example, one author's description of character development as requiring "the very hard work of feeling a character deeply, of understanding what happiness or hurt drives that character, so that it becomes impossible to lay superficial judgment on any behavior" (Silver 2010). Similarly, readers invoke characters in descriptions of their emotional responses to fiction, as in the teenage *Gatsby* enthusiast's powerful three-paragraph essay about the disgust, attraction, and exasperation that Daisy inspires in him. If fictional characters are abstracta, they are in no way appropriate objects of such attitudes; furthermore, being (presumably) causally inert, they can neither have, nor bring about, reactions of any kind (cf. Friend 2000: 999).

Some abstract object theorists maintain that authors do not write *about* fictional characters (see, for example, Searle 1975, van Inwagen 1977, and Salmon 1998). Nonetheless, on their view, a name that appears in a fictional work can be "imported" into reality, and used to make true assertions about a character—which, according to Salmon, is "the very same thing that it is the name of according to the story" (Salmon 1998: 80). If, however, the author's uses of the name in the novel do not refer to a character, it is not clear how, within that novel, properties are nonetheless ascribed to that character. (According to van Inwagen, the relation of ascription of a property to a character in a work of fiction "must be taken as primitive" (1979: 306).)

Finally, taking fictional names to refer to actually existing abstract objects appears to commit us to the falsity of sentences that we reflexively assent to, like (7):

(7) Daisy Buchanan does not exist.

To this, proponents of the view respond that (7) has different uses, and that, although it is true on some of those uses, it is not true on all of them. Someone might utter the

sentence to assert something that these theorists accept—for example, that *there is no person Daisy Buchanan*, that *Daisy Buchanan is not a real human*, or that *there is no object that actually possesses the properties ascribed to Daisy Buchanan in the novel*. Only when (7) is used to assert something like *there is no Daisy Buchanan*, or *the name 'Daisy Buchanan' does not refer* do these theorists object, since, on their view Daisy Buchanan (the fictional character) does exist. Thus, abstract objects theorists maintain that they can recognize true readings for (7); all that they deny is that (7) *must* be given a true reading.

Fictional Names Refer to Non-Existent Objects

Meinong (1904), however, takes (7) to be straightforwardly true. He draws a distinction between *being* and *existence*, and argues that an object need not exist in order to be the subject of thought and talk. A thing could, for example, be a male human who walks, talks, loves, and, in general, has the features attributed to Gatsby in the novel, while lacking the property of existence; that nonexistent object would be the referent of the name "Jay Gatsby," and (7) would be true in virtue of the fact that it does not exist.

Van Inwagen (1977: 299) gives a stark statement of one objection to this proposal: it appears to "necessitate an abandonment of what are commonly called 'the laws of logic'." That is: How can we consistently maintain that there *is* something that does not exist? Moreover, if the referent of a fictional name possesses all of the properties attributed to the character within the fiction, then we should expect the referent of "Gatsby" to possess the property of existence, since, in the novel, Gatsby is portrayed as an *existing* walking and talking man.

Parsons' (1980) account aims to overcome these challenges (Zalta 1983 and Zemach 1993 for alternative accounts of non-existent objects). He distinguishes between "nuclear" properties (like *red*, *round*, *sad*, *human*, etc.—intuitively, features by means of which an object could be identified), any set of which corresponds to an object, and "extra-nuclear" properties like existence. A fictional object, Parsons maintains, possesses only the nuclear properties attributed to it within the story. Thus, sentences (1) and (7) are true, while (2), (4), (5), and (6) are false; the truth-status of (3) is less clear. For the majority of philosophers, however, the question of whether we can arrive at a formally satisfactory account of the truth and falsehood of sentences by way of non-existent objects is academic, since they remain unreconciled to the supposed distinction between being and existence: "Is Gatsby a man? Yes. If he is pricked (or shot), does he not bleed? Yes. Is he worth more than the whole damn bunch put together? Probably. Just remember, though: he doesn't exist!"

Fictional Names Refer to Possible Objects

A third account of fictional names (developed in Lewis 1978) takes "Jay Gatsby" to refer to a man who really does exist, and who walks, talks, and loves—albeit as a denizen of some other (concrete) possible world. Straightforward readings of (7) come out false on this view; however, since "Jay Gatsby" doesn't refer in the actual world, (7) can be assessed as true if it is understood as the assertion that *Jay Gatsby is not actual*.

According to the possible objects theorist, at least some of the stories that are fictional in our world—including the one presented in *The Great Gatsby*—are, in some other possible worlds, "told as known fact". In each such possible world, the name "Jay Gatsby" refers to the individual who "plays the role" of Gatsby in that world.

An utterance of (1)—which, according to Lewis, abbreviates "In *The Great Gatsby*, Jay Gatsby lives in a lavish mansion in West Egg"—is true just in case in every world w in which the *Gatsby* story is told as known fact, the man who plays the role of Gatsby in w lives in a lavish mansion in West Egg. Thus, it is not really accurate to say that the possibilist picture provides *a* referent for the name "Jay Gatsby"; rather, it provides a (different) referent in each of the possible worlds determined by the story.

One advantage of taking fictional names to refer to possible objects is that it allows us to recognize the truth of assertions that go beyond what is explicitly predicated of characters within a fictional work. In all worlds in which the story is told as known fact, the woman who plays the role of Daisy will have a determinate number of freckles on her legs, and so (6) is true. (5) will be true, according to Lewis, just in case among the worlds in which the *Gatsby* story is told as known fact, a world in which Daisy has a larger wardrobe than Nick differs less from our actual world than any world in which Daisy does *not* have a larger wardrobe than Nick.

Challenges to the possibilist account of fictional reference fall into two categories. In the first are general concerns about modal realism: for example, that we seem to have no means by which to verify anything about the concrete possible worlds posited by Lewis, including their existence. The second variety of objection holds that, even if we grant that such possible worlds exist, we should nonetheless deny that the referents of our fictional names are to be found there. Lewis himself recognizes (1978: 263) that he gives no account of truth for sentences like (3) and (4), and that his proposal is in that respect incomplete. Furthermore, Donnellan (1970) and Kripke (1972) argue that only an object that is causally or historically connected to a name's introduction and subsequent use can be the referent of that name; if they are correct, then only objects in the actual world are candidate referents for names from fiction. Thomasson (2009: 16) finds the possibilist's identity conditions for fictional characters too restrictive, since they entail that if the author had ascribed to the character "so much as one different property (however trivial), she would have written about a different possible person." While this does seem like a dramatic consequence, the possibilist will presumably not be bothered by it, since, on his view, an author who ascribed a different property would *be* a different person: any scenario involving what did not actually happen is part of some other possible world.

Fiction and Make-Believe

As noted in the previous section, some theorists who hold that fictional names refer acknowledge that, when an author is writing her novel, "there is no point in debating what sort of thing" she is writing about, since she is not writing *about* anything (van Inwagen 1977: 301). But if we accept this, and if our spirits have been dampened by challenges to the referentialist views discussed in the previous section, we might well find ourselves reconsidering the enterprise of seeking referents for fictional names.

Bertrand Russell (1919: 169) warns of the folly of any such project; on his view, "it is of the very essence of fiction that only the thoughts, feelings, etc., [of an author] and his readers are real, and that there is not, in addition to them, an objective [character]." Russell takes those who posit referents for fictional names to have lost the "robust sense of reality" that is required for successful theorizing in this area. Some recent theorists who join Russell in denying that names from fiction refer do so with a robust sense of a particular part of reality: our games of make-believe.

These theorists (most prominently, Walton 1990; see also Currie 1990) characterize works of fiction as "props" in coordinated exercises of the imagination. Indeed, it is in virtue of playing a particular role in games of imaginative pretence, rather than any intentions on the part of an author, that a text is a work of fiction: "the notion of objects serving as props is independent of and has priority over that of the action of making objects to be used as props" (Walton 1991: 428). As with all varieties of make-believe, to consider which reports about fictional content are really true, or to attempt to externally assess the truth of what is asserted within a fiction, is to make a basic mistake. This becomes clearer when we consider two children who spend a balmy summer afternoon inhabiting the roles of Antarctic explorers adrift in iceberg-infested waters. Though the children may later tell family members lurid tales of their adventures, they would surely grow impatient with anyone who insisted on being told whether the explorers had *really* been growing weak from hunger, and whether they had *really* been able to eat their fill of plankton. On the one hand: "Yes, that's exactly what we were pretending!" And on the other: "We were just pretending—there aren't any such explorers!"

Walton's characterization of fictional 'truth' instead involves "there being a prescription or mandate in some context to imagine something" (1990: 39). When we follow the mandate, and imagine that what some sentence expresses is true, and that the names within it refer, we are not, however, imagining *of* something that it is a certain way. This is pretence, after all, and our pretendings take the form *there is a man named 'Jay Gatsby,' and he lives in a lavish mansion in a place called 'West Egg.'* Neither *The Great Gatsby* story, nor the imaginative uses to which we put it, make it the case that (1) is true, although (1), unlike (2), is fictionally true, or true in the story—just as, in the story, the name "Jay Gatsby" is a name of a man, while "Kengo Matsusaka" is not. Pretence theorists also deny that sentences like (3) and (4) are true, or that assertive utterances of them commit us to the existence of (real) fictional characters. Instead, they maintain that those who assert or assent to utterances of these sentences are engaged in a particular variety of make-believe, in which what is pretended is that there are fictional characters (Walton 1990, Brock 2002).

Walton's (1990) view is that, like fictional truth and 'real' truth, cognition and emotion within games of pretence should be distinguished from their nonfictional counterparts. Given that I 'believe' that Gatsby has just been shot, it is not surprising that I 'feel' sad and bitter. However, it is only within the pretence that I believe that Gatsby has been shot, and that I feel sad and bitter (that is: I am not experiencing *real* sadness and bitterness). What is crucial to notice, according to Walton, is that the 'feelings' (or quasi-emotions) that are elicited by fiction are not motivating: a moviegoer may, during a film about deadly green slime, tremble and clutch his chair with sweaty hands, but he does not run out of the theater, or believe that he needs to quickly move to higher ground, as he would if he were *really* afraid. In contrast, Lamarque (1981) argues that someone can be genuinely frightened by the thought of green slime, even if he is not frightened *that*, for example, the slime will soon engulf the theater.

Walton (1990) maintains that we are sometimes directed by fiction to imagine something that we actually believe, or know, to be true. According to New (1996), this position reveals problematic elements of Walton's portrayal of imagination and fiction. New maintains that someone who believes that New York is the most exciting city in the world, and encounters the sentence "New York is the most exciting city in the world" in a work of fiction will not *imagine*, or make-believe, that the sentence is true, or that the name "New York" refers. Rather, the reader's thought will involve the actual city

and its actual place at the pinnacle of exciting cities; thus, engagement with a work of fiction cannot, according to New, be understood as an exclusively imaginative exercise. (For a richly detailed discussion of mental pretence, see Camp 2009; for discussion of empirical results linking imagination and belief, see Nichols 2004.)

The pretence account also allows for the questionable possibility that a scrupulously accurate report of events that really did occur could qualify as a piece of fiction if it came to be used as a prop for certain imaginative exercises. The reverse is true as well: A work produced as fiction would cease to be such if it no longer figured in our practices of make-believe. Currie (1990) maintains that even if those who authored the Bible did so with fictional intentions—that is, even if the Bible was authored by those who intended it to serve as a prop in games of make-believe—the fact that "sufficiently many" people do not now read biblical stories as fiction means that the Bible could not, now, be properly classified as fiction. Thus, the pretence account seems to have the counter-intuitive consequence that being a work of fiction amounts to being *treated* as fiction. But this cannot be correct: We can, after all, ask about a text that is identified as a piece of fiction whether it in fact is, and we have a general idea of what sorts of investigations we might carry out in order to arrive at an answer.

Truth in Fiction

Our focus so far has mainly been on accounts of truth for individual sentences within, or about, fiction. But philosophers have also considered whether a fictional work, in its entirety, could express or present a truth, and have attended to the worry that it is not possible to *learn* such truths via fiction, since it would seem that even a work that expressed a truth could not provide evidence in support of it.

Critical assessments of a work of fiction often aim to make explicit the propositions that it expresses, or that it is committed to. One such analysis of *The Great Gatsby*, for example, takes the novel to provide "profound corrective insights into the nature of American experience," most notably, that "the American dream, stretched between a golden past and a golden future, is always betrayed by a desolate present" (Bewley 1954). Since no such sentence about the American dream appears in the novel, we need some account of how a text might present content that is not expressed by any particular sentence(s) within it.

On one view, some fictional works express philosophical *theses* concerning "man and his world," which are revealed to us via examinations of "plot, characters, dialogue, authorial interpolation, tone and themes" (Weitz 1975: 31). Exactly how, and whether, the plot and additional elements might be related to such theses is a matter of much dispute among literary critics. To the extent, however, that we offer arguments in support of the correctness of our particular interpretation of a novel's 'deeper meaning', or feel unease if we begin to think that what is 'said' by a favorite work of fiction might be objectionable, it is plausible to suggest that we do take a work to have, or be committed to, this sort of thesis (see Harcourt 2010).

Other theorists maintain that there is no content that fiction presents assertively, or 'as true.' Beardsley (1981) for example, holds that a work of fiction might well have *a* meaning discernable by an interpreter, but he denies that the work could advocate for (let alone establish) that meaning. In support of this position, Lamarque and Olsen (1994) emphasize that works of fiction rarely contain anything that could be construed as arguments for theses that they are purported to advocate.

At least some of the candidate truths from fiction are not, however, of the sort that we would expect to learn via arguments. Nagel (1974) and Jackson (1982) maintain that not all truths are propositional, or objectively accessible; some are essentially experiential, or subjective. These truths concern *what it is like*—for example, to be in a certain situation, to be part of a certain cultural group, or to be a certain kind of person—and, according to some philosophers, we can encounter such truths via fiction, and perhaps even receive confirmation of them.

Nussbaum (1985: 516) argues that we can develop our "moral perception" by actively engaging with fiction: "certain novels are, irreplaceably, works of moral philosophy." Reading such works with sufficient appreciation is a means of *becoming* sufficiently sensitive, and appropriately responsive, to particular complexities that might be present in human lives. Echoing Nussbaum, those who advocate for fiction's place within the curriculum for the helping professions argue that such practitioners must be capable of taking the perspective of those that they serve: to successfully counsel or advocate for the bereaved, for example, "the *feelings* associated with bereavement . . . must either be experienced first-hand, or else vicariously through writing or other media which *deliberately* attempt to invoke them, that is, which resonate with our own feelings" (Rolfe 2002: 91).

One might worry, however, about how we are to move from acknowledgment that fiction is a means of adopting a perspective other than our own, to any claim that such perspectives from fiction can provide *truth*, in ways that allow us to recognize it as such. The main source of difficulty here is that a work's effectiveness in moving the reader to take a particular perspective, and 'see' things from that perspective, seems in principle to be independent from the truth of what is viewed from that perspective.

As an example, consider Tolstoy's powerful short novel *The Kreutzer Sonata* (1889), in which the central character, Pozdnyshev, explains how and why he murdered his wife and the man that he suspected of being her lover. As his narrative unfolds, we become enlightened regarding the causes of his actions; he himself comes to understand that the love and hate that he felt for his wife ". . . were one and the same animal feeling only at opposite poles" (p. 126), and that "sexual passion, no matter how it is arranged, is evil, a terrifying evil . . ." (p. 110). (In an Afterward, Tolstoy elaborates on the lessons that his novella is meant to impart, among them, "that men and women [should] be educated in their families and through social opinion" so that "they would not regard falling in love and the carnal love connected with it as a poetic and elevated stage as they look on it now, but rather as a state of bestiality degrading for a human being . . ." (p. 164).)

Surely, for many readers, the perspective on offer in *The Kreutzer Sonata* is one that leaves them cold. Philosophers (Walton 1990, Gendler 2000, Weatherson 2004) have provided varied accounts of the phenomenon of imaginative resistance—that is, readerly refusal, or inability, to follow the imaginative prescriptions within a fiction. But not all readers will resist the invitation to take the perspective of Pozdnyshev, and to see things as he does. A skilled storyteller, writing with the fevered passion of true conviction, can make it seem that the world that he is giving us is, in fact, *the* world. And the perspectives to which we are provided access via fiction must always be subject to extra-fictional scrutiny; we cannot do without arguments, proceeding by way of objectively accessible propositions, if we are to assess purported truth in fiction.

Bibliography

Beardsley, M. (1981) *Aesthetics: Problems in the Theory of Criticism*, 2nd edition, Indianapolis: Hackett.

Bewley, M. (1954) "Scott Fitzgerald's Criticism of America," *The Sewanee Review* LXII (Spring).

Braun, D. (2005) "Empty Names, Fictional Names, Mythical Names," *Noûs* 39, pp. 596–631.

Brock, S. (2002) "Fictionalism about Fictional Characters," *Nous* 36, no. 1, pp. 1–21.

Byron, G. (1824) *Don Juan*, Cantos XII.–XIII.–-and XIV. New York: Charles Wiley.

Camp, E. (2009) "Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments," *Midwest Studies in Philosophy* XXXIII.

Chomsky, N. (1988) *Language and Problems of Knowledge: The Managua Lectures*, Cambridge, MA: The MIT Press.

Currie, G. (1990) *The Nature of Fiction*, Cambridge: Cambridge University Press.

Donnellan, K. (1970) "Proper Names and Identifying Descriptions," *Synthese* 21, pp. 335–358.

Fitzgerald, F.S. (1925) *The Great Gatsby*, New York: Scribner's.

Friend, S. (2000) Review of Thomasson (1999), *Mind* 109, pp. 997–1000.

Gendler, T. (2000) "The Puzzle of Imaginative Resistance," *The Journal of Philosophy* 97, no. 2, pp. 55–81.

Harcourt, E. (2010) "Truth and the 'Work' of Literary Fiction," *British Journal of Aesthetics* 50, no. 1, pp. 93–97

Jackson, F. (1982) "Epiphenomenal Qualia,"*Philosophical Quarterly* 32: pp. 127–136.

Kripke, S. (1972) *Naming and Necessity*, Oxford: Blackwell.

Lamarque. P. (1981) "How Can We Fear and Pity Fictions?" *British Journal of Aesthetics* 21, pp. 291–304.

Lamarque, P. and Olsen, S. H. (1994) *Truth, Fiction, and Literature: A Philosophical Perspective*, Oxford: Clarendon.

Lewis, D. (1978) "Truth in Fiction," *American Philosophical Quarterly* 15, pp. 37–46. In *Philosophical Papers*, vol. 1 (1983), Oxford: Oxford University Press.

Meinong, A. (1904) "On the Theory of Objects," in R. Chisholm (ed.) *Realism and the Background of Phenomenology* (1960) Atascadero: Ridgeview.

Nagel, T. (1974) "What Is It Like to Be a Bat?" *Philosophical Review* 83, pp. 435–50.

New, C. (1996) "Walton on Imagination, Belief and Fiction," *British Journal of Aesthetics* 36 no. 2, pp. 159–165.

Nichols, S. (2004) "Imagining and Believing: The Promise of a Single Code," *The Journal of Aesthetics and Art Criticism*, Special Issue on Art, Mind, and Cognitive Science.

Nussbaum, M. (1985) " 'Finely Aware and Richly Responsible': Moral Attention and the Moral Task of Literature," *The Journal of Philosophy* 82, no. 10.

Parsons, T. (1980) *Non-Existent Objects*, New Haven: Yale University Press.

Rolfe, G. (2002) " 'A Lie That Helps us See the Truth': Research, Truth and Fiction in the Helping Professions," *Reflective Practice* 3, pp. 89–102.

Russell, B. (1919) *Introduction to Mathematical Philosophy*, London: Allen and Unwin.

Salmon, N. (1998) "Nonexistence," *Noûs* 32, no. 3, 277–319.

Searle, J. (1975) "The Logical Status of Fictional Discourse," *New Literary History* 6, No. 2, pp. 319–332.

Silver, M. (2010) "The Rules of the Game," posted on *The Elegant Variation* weblog, http://marksarvas.blogs.com/elegvar/2010/04/the-rules-of-the-game.html

Thomasson, A. (1999) *Fiction and Metaphysics*, Cambridge: Cambridge University Press.

—— (2003) "Speaking of Fictional Characters," *Dialectica* vol. 57, no. 2, pp. 207–226.

—— (2009) "Fictional Entities," in J. Kim and E. Sosa (eds.) *A Companion to Metaphysics*, Blackwell, pp. 10–17.

Tolstoy, L. (1889) *The Kreutzer Sonata* (1890) "Epilogue to *The Kreutzer Sonata*," reprinted in *Six Short Masterpieces by Tolstoy*, translated by M.Wettlin, New York: Dell Publishing Co.

van Inwagen, P. (1977) "Creatures of Fiction," *American Philosophical Quarterly* 14, no. 4, pp. 299–308.

Walton, K. (1990) *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press.

—— (1991) "Reply to Reviewers," *Philosophy and Phenomenological Research* 51, no. 2, pp. 413–431.

Weatherson, B. (2004) "Morality, Fiction, and Possibility," *Philosophers' Imprint* 4, no. 3.

Weitz, M. (1975) "Literature without Philosophy: 'Antony and Cleopatra'," in K. Muir (ed.), *Shakespeare Survey*, Cambridge: Cambridge University Press, pp. 29–36.

Wolterstorff, N. (1980) *Works and Worlds of Art*, Oxford: Clarendon Press.

Yagisawa, T. (2001) "Against Creationism in Fiction," *Philosophical Perspectives* 15, pp. 153–172.

Zalta, E. (1983) *Abstract Objects*, Dordrecht: Reidel.

Zemach, E. (1993) "Existence and Nonexistents," *Erkenntnis* 39, pp. 145–166.

—— (2003) Review of Thomasson (1999), *The Philosophical Review* 112, no. 3, pp. 427–431.

Section II

FOUNDATIONS OF SEMANTICS

2.1

REFERENCE

Teresa Robertson

1 Introduction: Reference, Semantic Content, and the View that Content is Reference

Reference may be the most basic of all semantic relations. It is the relation that an expression (in a particular context) bears to the thing (or things) for which it stands. (From here on, reference to contexts is suppressed unless suppression is apt to cause confusion.) The clearest examples of this relation in natural language are those involving the bare *demonstratives* 'this' and 'that'. When one says, gesturing to the coffee in one's cup, "This is good and strong," the expression 'this' refers to the coffee in the cup. Similarly, when I say, indicating Seattle, "That's my hometown," the expression 'that' refers to Seattle. Here are some other natural (which is not to say uncontroversial) examples of the reference relation: The *proper name* 'Socrates' *refers to* (or in other terminology *designates* or *denotes*) the philosopher Socrates; the *definite description* 'the philosopher executed for corrupting the youth of Athens' refers to him too; and the pronoun 'I' (in a suitable context) refers to me. The expressions in the five examples just given are all naturally thought of as *singular terms*, terms whose fundamental semantic role is to refer to a single individual. Paradigmatic singular terms are the individual constants and variables of formal logic. Because the notion of reference finds its natural home in the relation between expressions for single individuals and the individuals themselves, the bulk of the philosophical work on reference has been concerned with singular terms or expressions that are at least arguably singular terms.

Interest in reference has been intimately bound up with interest in meaning, specifically with interest in *semantic content*. The philosopher's notion of semantic content is a thoroughly intuitive one. (Again, something may be intuitive or natural or naive but nonetheless controversial.) Consider the English sentence 'Snow is white' and the French sentence 'La neige est blanche'. Intuitively, these sentences say the same thing. The single thing that both sentences say (or *mean* or *express*) is the *proposition* that snow is white. It is this proposition that is said to be the semantic content (or simply the *content*) of each of these declarative sentences. It is natural to say that what those sentences express is true. Taking appearances at face-value, propositions are taken to be bearers of truth-value (that is, things that are true or false). For similar reasons, they are taken to be bearers of modal properties (like necessity and contingency) and epistemic properties (like being knowable a priori or being knowable only empirically). They are also taken to be objects of other attitudes (that is, in addition to being taken to be objects of knowledge, they are taken to be objects of assertion, belief, doubt, and the like).

Propositions are commonly thought of as structured entities with component parts. Consider the propositions expressed by the following sentences: 'Socrates is wise', 'Socrates is Greek', and 'Plato is Greek'. Intuitively the proposition expressed by the second sentence shares a component with the proposition expressed by the first sentence—a component that is associated with the word 'Socrates'—in virtue of which each proposition is concerned with the individual Socrates. Similarly, the proposition expressed by the second sentence intuitively shares a component with the proposition expressed by the third sentence—a component that is associated with the phrase 'is Greek'—in virtue of which each proposition is concerned with the property of being Greek. Consideration of the propositions expressed by slightly more complex sentences like 'Socrates taught Plato' and 'Plato taught Socrates' reveals that propositions must be structured entities: if they were not, then the two propositions, which are clearly different since one is true and the other false, would not be distinguishable since both would simply consist of the same components. It is natural to think of the structure of the proposition as at least roughly isomorphic to the structure of the sentence that expresses it.

The idea that the content of a compound expression is a function of the contents of its parts together with the way those parts are structured is nearly irresistible—at least as a general rule. There are numerous exceptions to this rule. For example, whatever the content of 'Socrates' is—that is, whatever 'Socrates' typically contributes to the content of sentences in which it occurs—that thing is evidently no part of the following sentence:

'Socrates' contains eight letters.

Nonetheless it is safe to assume, as most work in semantics does, that this rule holds by and large. I will take it for granted in this chapter. In fact, for simplicity, I will typically assume something stronger, namely that by and large the content of a compound expression is a structured entity composed of the contents of the expression's contentful parts. I will say that an expression is *compositional* when its content is a function of the contents of its parts, together with the way those parts are put together (or, more strongly, when its content is a structured entity composed of the contents of the parts).

The most straightforward theory about content says simply that content *is* reference. On this theory, the proposition that is the content (and referent) of the sentence 'Socrates is wise' has as its components the individual Socrates (which is the referent of 'Socrates') and the property of being wise (which can, at least for present purposes, be taken to be the referent of 'is wise'). This theory unites the semantic relations of referring and expressing.

2 Trouble over Definite Descriptions

Two of the naive ideas encountered in §1, namely that definite descriptions are singular terms and that content is reference, come into conflict with one another. Consider again the definite description 'the philosopher executed for corrupting the youth of Athens'. Intuitively, it has contentful parts—parts which themselves have semantic content. And intuitively descriptions are compositional. (Again, exceptions may arise. Consider the description 'the number of letters in "Socrates"'. Whatever the content of 'Socrates' is, it is no part of the content of 'the number of letters in "Socrates"'. But here it is the quotation marks and not some feature common to all definite descriptions

that is responsible for the exception.) Accepting these intuitions at face-value reveals that we cannot accept both the straightforward semantic theory that content is reference and the natural thought that 'the philosopher executed for corrupting the youth of Athens' is a singular term that refers to Socrates: According to modern anatomy at least, Socrates is not himself a complex entity composed of the semantic contents of the contentful parts of 'the philosopher executed for corrupting the youth of Athens'. There are two obvious ways to deal with this problem, and historically both have had advocates: Bertrand Russell (Russell 1905) rejected the claim that definite descriptions are singular terms and John Stuart Mill (Mill 1893) rejected the claim that content is reference. I will sketch both of these kinds of responses, though I will not aim to discuss the particular views of Russell and Mill themselves. I will use 'Russell$_1$*' and 'Mill*' as labels for the views I discuss.

3 Definite Descriptions: Russell$_1$* and Mill*

Russell$_1$* resolves the trouble over definite descriptions by retaining the thought that content is reference but giving up the thought that definite descriptions are singular terms. Russell$_1$* assimilates definite descriptions to other determiner phrases (phrases of the form 'every F', 'some Fs', 'most Fs', etc.), which are viewed as *restricted quantifier phrases*. So far, Russell agrees with Russell$_1$*. Russell himself thought that such phrases did not have content "in isolation"; however on a slightly more natural view, a restricted quantifier's content is a higher-order property—perhaps a property of properties or a property of classes. For example, the sentence 'Every philosopher is Greek' may be thought of as attributing to the property of being Greek the property of being (a property that is) instantiated by everything that instantiates the property of being a philosopher, so that the attributed property (of properties) may be thought of as the content of 'every philosopher'. (Roughly this idea is due to Richard Montague (Montague 1974).) Similarly, a sentence like 'The philosopher executed for corrupting the youth of Athens was Greek' may be thought of as attributing to the property of being Greek the property of being instantiated by something that uniquely instantiates the property of being a philosopher executed for corrupting the youth of Athens.

In contrast, Mill* resolves the trouble over definite descriptions by retaining the thought that definite descriptions are singular terms but giving up the thought that content is reference. The idea is that there are (at least) two kinds of semantic value—content and reference—and that the content of a definite description is not its referent but instead some complex entity composed of the semantic contents of the description's contentful parts. On this view, the description 'the philosopher executed for corrupting the youth of Athens' has as content some complex involving such things as the property of being a philosopher, the property of being executed, and so on, but the view retains the natural thought that the description's referent is Socrates. Deviating as little as necessary from the naive theory that content is reference, Mill* allows that some expressions, and in particular proper names, have as their contents their referents. (It is worth stressing that one can recognize a thoroughgoing distinction between content and referent, as Mill* does, while allowing that for some expressions the very same entity is both. Compare: there is a thoroughgoing distinction between sister and wife, although for Zeus the same entity is both.)

4 Challenges for Russell$_1$* and Mill*

Though the theories of Russell$_1$* and Mill* are very different in their theoretical outlooks, with the former but not the latter accepting that content is reference, they share a commitment to the claim that the semantic content of a proper name is its referent. This claim faces challenges in the form of various puzzles that can be grouped into two kinds: puzzles involving substitution of coreferential names and puzzles involving empty names (that is, names that lack referents).

Let's first look at the substitution puzzles. My father's given name is 'Charles', but for reasons that are lost to time, people in my family usually call him 'Pete'. Now consider the following two sentences.

(1) Charles is Charles.
(2) Pete is Charles.

Assuming that these sentences are compositional, the theories of Russell$_1$* and Mill* ascribe the same content to both sentences, because they ascribe the same content to 'Charles' and 'Pete'. Or, more fully, because the components of the proposition expressed by each sentence are the same (my father, the relation of identity, and my father again) and they are put together in the same way. But, the sentences do not seem to express the same proposition, since the proposition expressed by (1) is not informative whereas the proposition expressed by (2) is. This is the *puzzle of informativeness* presented by Gottlob Frege (Frege 1892). He presented the puzzle using identity sentences, as I have done, but that is not crucial to the puzzle, as the following sentences reveal.

(3) Charles is a person, if Charles is a person.
(4) Pete is a person, if Charles is a person.

Again the theories of Russell$_1$* and Mill* ascribe the same content to both sentences, yet intuition tells us that (3) and (4) express different propositions—one that is not informative and one that is.

There are a number of similar puzzles that vary from Frege's only in what particular property of propositions they highlight. So, for example, one can generate a puzzle concerning epistemic status (rather than informativeness) by noting that intuitively the proposition expressed by (3) is knowable a priori, whereas the proposition expressed by (4) is not. Similarly one can generate a puzzle concerning belief by noting that although Clark Kent is Superman, Lois believes the proposition expressed by (5) but not the one expressed by (6).

(5) Clark Kent is a reporter.
(6) Superman is a reporter.

(Here I am treating the comic book fiction as fact.) (A family of closely related substitution puzzles involves sentences of the form ⌜It is informative that S⌝, ⌜It is knowable a priori that S⌝, ⌜So and so believes that S⌝, and so on, where the numbered sentences above are examples of substituends for 'S'. Because these sentences involve the operator 'that', the full statement of these puzzles would have to address issues that are beyond the scope of this chapter. In brief, 'that' appears to operate somewhat as quotation marks

do, so that it takes significantly more than compositionality to underwrite the claims that according to Russell$_1$* and Mill*, 'Lois believes that Clark Kent is a reporter' and 'Lois believes that Superman is a reporter' express the same proposition. Nor does compositionality provide enough to underwrite the claim that according to Russell$_1$* and Mill* those two sentences have the same truth-value. Those claims can be adequately supported, but it takes additional theory to do so.)

It is important to understand that the substitution puzzles that arise for Russell$_1$* and Mill* do not arise when definite descriptions are involved. These theories are committed to saying that coreferential names have the same semantic content, but they are not committed to saying that coreferential descriptions have the same semantic content or to saying that a name and a coreferential description do. (Here I use "referential" in a loose way, a way according to which even Russell$_1$*—who strictly speaking thinks that the referent of a definite description is a higher-order property rather than the individual that is intuitively its referent—can say that 'the philosopher executed for corrupting the youth of Athens' and 'Socrates' are coreferential.) And this allows these theories to resist the claims that sentences like (7), (8), and (9) below have the same content, even though Benjamin Franklin was the first postmaster general and also the inventor of bifocals.

(7) Benjamin Franklin was interested in science.
(8) The first postmaster general was interested in science.
(9) The inventor of bifocals was interested in science.

This means that no analogs of the substitution puzzles involving names arise in the case of definite descriptions.

Now, let's look at the most general of the puzzles involving empty names. Intuitively, the name 'Santa' has no referent. Nonetheless, sentences involving it appear to be contentful. Consider the following sentence.

(10) Santa is jolly.

What is the content of this sentence according to Russell$_1$* and Mill*? An entity composed of the referent of 'Santa' and the property of being jolly. But 'Santa' does not have a referent. So how can (10) be contentful? One natural response to this question is that the proposition expressed by (10) must be some kind of "gappy" structured entity, with a gap where there should be something to which 'Santa' refers. But this solution founders when we consider the following sentence.

(11) Tinkerbell is jolly.

Assuming that 'Tinkerbell' has no referent, (11) must also express a gappy proposition; in fact the same one expressed by (10). But intuitively (10) and (11) do not express the same proposition. (There are other puzzles involving empty names that are less general than this one. The most famous of these is the *puzzle of true negative existential statements*, which involve sentences like 'Santa does not exist'. A full discussion of this and the other more specific puzzles involving empty names is beyond the scope of this chapter.)

Again, it is important to understand that for Russell$_1$* and Mill* there is no analog to the general puzzle involving empty names that involves instead empty descriptions.

For both Russell$_1$* and Mill* a description like 'the largest prime' has content: for Russell$_1$*, its content is something like the property of being instantiated by something that uniquely instantiates the property of being a prime greater than all others; for Mill*, its content is some complex involving the semantic contents of the semantically contentful constituents of the description.

5 Addressing the Challenges for Russell$_1$* and Mill*: Russell$_2$* and Frege*

Since the puzzles that arise for Russell$_1$* and Mill* do not arise when descriptions are involved, the puzzles can be solved by assimilating names to descriptions. This is accomplished in a way that retains much of the theoretical outlook of Russell$_1$* (in particular adherence to the simple theory that content is reference) by Russell$_2$*, who maintains that names are really just definite descriptions in disguise (Russell 1910–1911). (Again, my use of a "starred name" indicates that my aim is not to capture the view of a particular historical figure so much as to tell a purely philosophical story in which figures a view that bears important similarities to the view of the referent of the "unstarred name.") On this view, a name like 'Santa' is semantically equivalent to a description like 'the fat jolly old elf who delivers presents to good children on Christmas', which in turn is thought of as a restricted quantifier phrase. The assimilation of names to descriptions is accomplished in a way that retains much of the theoretical outlook of Mill* by Frege*, who takes Mill*'s recognition of a distinction between content and referent to its extreme, holding that for every expression, that which serves as content includes a mode of presentation of that which serves as referent (Frege 1892). The content is thus distinct from the referent, at least typically. On this view, names, like descriptions, have contents that are conceptual in nature. The name 'Santa' has as its semantic content some complex involving such things as being fat, being jolly, and so on, so that 'Santa' is semantically equivalent to some definite description like 'the fat jolly old elf who delivers presents to good children on Christmas'.

6 Trouble for Russell$_2$* and Frege*

Though the theories of Russell$_2$* and Frege* provide solutions to the puzzles encountered in §4, there are three powerful arguments against these theories: the *modal argument*, the *epistemic argument*, and the *semantic argument*. (The first two arguments are due to Saul Kripke (Kripke 1972/1980). The last is due to Keith Donnellan (Donnellan 1972) and independently to Kripke.) Somewhat surprisingly, two of these arguments (the modal and epistemic) depend on considerations that are remarkably similar to those involved in the puzzles of §4.

Let's suppose with Russell$_2$* and Frege* that the content of 'Gödel' is the same as the content of some description, say 'the discoverer of the incompleteness of arithmetic'. Now consider the following two sentences.

(12) Gödel (if he exists) discovered the incompleteness of arithmetic.
(13) The discoverer of the incompleteness of arithmetic (if there is exactly one) discovered the incompleteness of arithmetic.

The modal argument points out that if 'Gödel' and 'the discoverer of the incompleteness

of arithmetic' have the same content, then (12) and (13) express the same proposition. (The argument assumes that these sentences are compositional.) But, the argument continues, the proposition expressed by (12) is not necessary, whereas the proposition expressed by (13) is. Hence (12) and (13) do not express the same proposition. Hence 'Gödel' and 'the discoverer of the incompleteness of arithmetic' do not have the same content.

The epistemic argument is very similar to the modal argument. Both arguments have the same general form and reach the same conclusion. In place of the modal argument's appeal to the fact that the proposition expressed by (12) is not necessary whereas the proposition expressed by (13) is, the epistemic argument appeals to the fact that the proposition expressed by (12) is not knowable a priori, whereas the proposition expressed by (13) is.

The semantic argument, which is sometimes confused with the modal argument, is a very different sort of argument from the modal and epistemic arguments. Suppose that in fact the person who we think discovered the incompleteness of arithmetic (the man we call 'Gödel') did not in fact discover the incompleteness of arithmetic but rather stole the result from his neighbor (a man whom we call 'Schmidt'). Who would the name 'Gödel' refer to—the thief or the neighbor? Clearly, it would refer to the thief. (Or, not to put too fine a point on it, 'Gödel' would still refer to *Gödel* and not to Schmidt.) Who would 'the discoverer of the incompleteness of arithmetic' refer to—the thief or the neighbor? Clearly, it would refer to the neighbor (that is, Schmidt) even though we would in the circumstances mistakenly think that it referred to Gödel. Thus, assuming that content determines referent, the content of the name 'Gödel' is not the same as the content of 'the discoverer of the incompleteness of arithmetic'.

Given these problems for Russell$_2$*-Frege* and the problems for Russell$_1$*-Mill* that were discussed earlier, it is natural to search for some third general view concerning the semantic content of proper names. But it is evident that there is precious little room between the view that the semantic content of a proper name is its referent and the view that the semantic content of a proper name is the kind of thing that could be expressed by some definite description. Nonetheless philosophers have attempted to skirt the problems facing the dominant views by claiming that a proper name has as its content a "nondescriptive sense". (An early advocate of this approach was Felicia Ackerman (Ackerman 1979).) Such a sense would have to be some sort of conceptual entity that is associated with the name and that is such that the thing that falls under the relevant concept would be the referent of the name, but such a sense would at the same time have to be something that could not be captured by any definite description. (A conceptual content that is not but that could be captured by a definite description would not be a nondescriptive sense.) The exact nature of this line of thought remains murky and consequently attractive to some. Philosophers have also tried more straightforward approaches. There are current theories in the spirit of Russell$_1$*-Mill* that address the problems of §4. (Prominent defenders of this line include Nathan Salmon and Scott Soames. Soames (2002) modifies Salmon's (1986) solution to the problems of substitution, and endorses Salmon's (1998) solution to the problems of empty names.) And also there are numerous theories in the spirit of Russell$_2$*-Frege* that attempt to solve the problems of §6. (Michael Dummett (Dummett 1973) and Gareth Evans (Evans 1982) gave early and influential responses to the arguments of Kripke and Donnellan.) No consensus has emerged.

7 Mechanisms of Reference

It has been thought that one advantage of the Russell$_2$*-Frege* view over the Russell$_1$*-Mill* view is that the former has something to say about the mechanism by which reference is secured while the latter stands in need of supplementation. To fill this need, Kripke (1972, 1980) proposed the so-called causal-historical picture of reference. On this picture, a name may be introduced in a variety of ways. (Here are just a few examples. Parents of a newborn may introduce a name for their child by deciding on a name and acting on that decision. Similarly, the adoptive parents of an alien baby with extraordinary superhuman powers may introduce a name for their child by saying, "Let's call him 'Clark'." A scientist may introduce a name for a particular individual by saying, "I'll use 'Eve' to name the most recent female ancestor common to all living humans." The ritual aspects of these dubbings are presumably brought to a bare minimum when, for example, someone who lives alone names her pet, but even workers at a law office may without ceremony simply begin using the name 'Romeo' to refer to whoever sent flowers to the managing partner.) Those who introduce the name use it to refer to the relevant object. In communication, the name gets passed on to other users, so that the name refers to that same object even in the idiolects of the newer users. Spelling out the exact details of the nature of the causal chains that link later uses to initial uses is a difficult empirical task. (For example, why is it that although there are presumably causal chains linking my use of 'St. Nicholas' as well as my use of 'Santa Claus' to early uses of a name of Nicholas of Myra, 'St. Nicholas' refers to that saint while 'Santa Claus' does not?) What is of philosophical interest is that on this picture, in sharp contrast to the Russell$_2$*-Frege* view, reference is determined contextually rather than conceptually.

While the causal-historical picture is just a picture and not a full-fledged theory (since many empirical questions about the nature of the relevant causal chains remain unanswered), the picture does provide enough of a supplement to the Russell$_1$*-Mill* view that it would be unfair to say that it remains a complete mystery on this view how a name comes to be hooked up to its referent. And in spite of what is often said about the Russell$_2$*-Frege* view, there is in fact some mystery on this view about this issue. While it is easy enough for the view to explain how a name's associated description gets connected to the name's referent (since the referent is simply whatever satisfies the description), the view stands in need of supplementation to explain how the name gets connected to its associated description.

8 Reference and General Terms

Although the reference relation is most at home in connection with singular terms, it is not much of a stretch to see the relation as holding between general terms and kinds. For example, it is natural to think of the general terms 'water' and 'tiger' as referring to a particular liquid substance and a particular species of animal respectively. (It is more of a stretch to see what the analog of the notion of reference is for other expressions such as predicates and sentences. An ingenious argument, the so-called slingshot, due to Alonzo Church (Church 1943) and Kurt Gödel (Gödel 1944), gives strong reason to think that the closest analog to singular term reference for a predicate is the class of (*n*-tuples of) individuals to which it applies and that the closest analog for a sentence is its truth value.) And in fact many of the issues that arose in the discussion of singular

terms arise in much the same way for general terms. In particular, the modal, epistemic, and semantic arguments against the view that a proper name has the same content as a singular definite description apply equally against the view that a simple general term, like 'water' or 'tiger', has the same content as a general definite description, like 'the colorless, odorless, thirst-quenching liquid that fills the lakes and streams' or 'the species that serves as mascot for Princeton University'. (I leave the details as an exercise for the reader.)

Though much of the literature on general terms focuses on one-word general terms for natural kinds, it is far from clear that such terms differ in semantically significant ways from one-word general terms for nonnatural kinds. In particular, the modal, epistemic, and semantic arguments can be run using, for example, 'crayon' and 'the favorite writing instrument of Yoko'. (Hillary Putnam (Putnam 1975) using his Twin-Earth thought experiment gave a version of the semantic argument in connection with natural-kind terms. He also argued that artifact terms, like 'pencil', are not semantically equivalent to any definite description.) This is not to deny that a one-word general term may be synonymous with a compound common noun phrase: surely 'bachelor' is synonymous with 'unmarried man'. This highlights a significant point of disanalogy between terms for kinds and terms for individuals: It is commonplace to refer to kinds not only with simple terms and definite descriptions but also with compound common noun phrases like 'unmarried man' and 'guy who won't take no for an answer'. This fact raises many questions. Given that 'bachelor' and 'unmarried man' are synonyms, should the considerations that favor a Russell$_1$*-Mill* view for simple-kind terms like 'bachelor' lead us to favor a Russell$_1$*-Mill* view for compound-kind terms like 'unmarried man'? Or rather does the apparent fact that 'unmarried man' has descriptive content give us a consideration in favor a Russell$_2$*-Frege* view for 'bachelor'? What impact should answers to these questions have on views about singular terms? Are there singular term analogs to kind-phrases? Suffice it to say that a more thorough understanding of workings of both singular and general terms is needed to get to the bottom of the notion of reference.

References

Ackerman, F., "Proper Names, Propositional Attitudes, and Nondescriptive Connotations," *Philosophical Studies* (1979), 55–69.

Church, A., "Review of Carnap's *Introduction to Semantics*," *Philosophical Review* 52 (1943), 298–304.

Donnellan, K., "Proper Names and Identifying Descriptions," in D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (Dordrecht: D. Reidel, 1972), 356–379.

Dummett, M., *Frege: Philosophy of Language* (New York: Harper and Row, 1973).

Evans, G., *Varieties of Reference* (Oxford: Oxford University Press, 1982).

Frege, G., "*Über Sinn und Bedeutung*," *Zeitschrift für Philosophie und Philosophische Kritik* 100 (1892), 25–50. English translation in M. Beaney (ed.), *The Frege Reader* (Oxford: Blackwell, 1997), 151–171.

Gödel, K., "Russell's Mathematical Logic," in P.A. Schilpp (ed.) *The Philosophy of Bertrand Russell* (New York: Tudor, 1944), 123–154.

Kripke, S., *Naming and Necessity* (Harvard University Press and Blackwell, 1972/1980).

Mill, J.S., "Of Names," in Book I, Chapter II of *A System of Logic* (New York: Harper and Brothers, 1893), 29–44.

Montague, R., "The Proper Treatment of Quantification in Ordinary English," in J. Hintikka, J. Moravcsik, and P. Suppes (eds.), *Approaches to Natural Language* (Dordrecht: Reidel, 1973), 221–242. Reprinted in Montague (1974), 247–270.

Montague, R., *Formal Philosophy: Selected Papers of Richard Montague*, edited and with an introduction by R. Thomason (New Haven: Yale University Press, 1974).

Putnam, H., "The Meaning of 'Meaning'," in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science VII: Language, Mind, and Knowledge* (Minneapolis: University of Minnesota Press, 1975), 130–193.
Russell, B., "On Denoting," *Mind* 14 (1905), 479–493.
Russell, B., "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotelian Society* 11 (1910–1911), 108–128.
Salmon, N., *Frege's Puzzle*, (Cambridge, MA: MIT Press, 1986).
Salmon, N., "Nonexistence," *Noûs* 32 (1998), 277–319.
Soames, S., *Beyond Rigidity: The Unfinished Semantic Agenda of* Naming and Necessity, (Oxford: Oxford University Press, 2002).

2.2

THEORIES OF TRUTH

Matti Eklund

Questions about truth are central to several different parts of philosophy. In what follows I will summarize central parts of the debates over truth in metaphysics, in philosophy of language, and, more briefly, in philosophy of logic. Throughout I will also be concerned with the relations between these different debates.

The Metaphysics of Truth

A preliminary question for all discussions of truth is what the primary bearers of the properties of truth and falsity are: are they propositions or sentences, or something else altogether, like acts of judgment? For the purposes of the first part of this discussion, I will assume that *propositions* are the primary truth-bearers.

The first question I will actually pause on is one which arguably underlies the debate over the traditional main theories of truth: the correspondence theory, the coherence theory, the identity theory, and deflationism. *By virtue of what* are propositions true and false, respectively? This is a question concerning the *relationality* of truth. We can distinguish four main types of answer: (a) propositions are true or false by virtue of how things stand in extra-propositional reality (except when they are about other propositions); (b) propositions are true or false by virtue of relations they stand in to other propositions; (c) the truth or falsity of a proposition is determined by intrinsic features; (d) the whole question of what a proposition's truth depends on is misbegotten (Wright 1999).

Alternative (a) should seem the most plausible. Even without getting into details regarding what a specific theory on which (b) is correct might look like, one might well think that when a proposition concerns how things stand in nonpropositional reality, its standing in the right relations to other propositions is not sufficient to make it true. Explanation (c) obviously runs into trouble with contingency: If the truth of a proposition is an intrinsic feature, how can a proposition have different truth-values with respect to different possible scenarios? I will turn to the dismissive suggestion (d) below.

Alternatives (a) through (d) correspond roughly to the important traditional theories of truth: (a) spells out the *correspondence* idea; (b), when supplemented by the claim that the relation in question is *coherence*, spells out the idea of coherence; (c) suggests the *identity theory*, on which true propositions do not correspond to but simply are facts (but see below); and (d) should be recognizable as something *deflationists* about truth are apt to hold.

There are ways of conceiving of a correspondence theory of truth, given which it does not seem as plausible as it does if we take the view to be captured by (a). A correspondence theory is often formulated as the thesis that truth consists in correspondence with *the facts*, and to be committed to an *ontology of facts*. This gives rise to a number of problematic questions—for example, concerning the existence of negative and hypothetical facts. However, one can say, consistently with (a), that what makes a given proposition true: say, that *a* is not F means simply that *a* fails to be F. There is no *obvious* need to appeal to such entities as facts. Second, a correspondence theory is sometimes claimed to be bound up with *metaphysical realism*. But (a) is not so bound up. One can say that a proposition is true by virtue of how things stand in nonpropositional reality, while holding that nonpropositional reality is mind-dependent in a way that is incompatible with realism. Third, some are skeptical of a correspondence theory on the grounds that it is hard to elucidate the notion of correspondence involved. It may be tempting to understand correspondence as having something to do with similarity. But can propositions really, in any interesting sense, be *similar* to facts? And if we are not to elucidate the relation of correspondence by saying that it is a matter of similarity, just how can correspondence be elucidated? However, it must be said that (a) can be true even if nothing more informative can be said. Fourth, in light of the above points, and in light of the fact that (a) doesn't by itself provide anything like a reductive characterization if truth, it may be thought that (a) by itself is too modest to count as a *correspondence theory*. But, importantly, (a) does seem to be in conflict with natural ways of spelling out other conceptions of truth, and (a) goes beyond some truly modest understandings of correspondence: for example those in which a correlation in truth of *the proposition that p is true* and *p* is sufficient for correspondence. Answer (a) adds to this that the truth of the proposition *depends on* what holds in nonpropositional reality.

Now turn to (b). The coherence theory, according to which truth is coherence with other propositions, is clearly one way to spell out (b). The coherence theory faces an immediate objection, known as the Bishop Stubbs objection, from Bertrand Russell's example: any proposition—including obviously false ones like the proposition that Bishop Stubbs was hanged for murder—is part of some coherent collection of propositions, but coherence with other propositions is not sufficient for truth (Russell 1910). The coherence theorist's response is to insist that truth consists in coherence with a privileged collection of propositions. The question is whether that is a workable way out. In his study of coherence theories of truth, Ralph Walker (1989) has prominently argued that there is still a problem: what accounts for the truth of the proposition that a given collection of propositions is privileged?

Coherence theory is often presented as an historically important competitor to correspondence theory. But the following complication must be stressed. Those traditionally counted as coherence theorists of truth—prominently the British idealists (for example, Joachim (1906) and Bradley (1909))—took reality itself to be an *organic whole* and so to be characterized by its coherence. Given this view on reality one can emphasize coherence without being a coherence theorist of truth in the sense of (b). For example, one can use this view on reality to deny that individual propositions are ever (fully) true and insist instead that only a system of propositions displaying a certain coherence can be true. But even that wouldn't be to embrace (b). It is only that proper correspondence obtains only between larger units than it is otherwise conceived of as obtaining between, and the units are themselves characterized by coherence. Another characteristic view

of the idealists was, precisely, *idealism*, the view that reality is mind-dependent. But it is possible to insist that reality is somehow mind-dependent, but that a proposition and the reality which makes it true are two distinct elements of this mind-dependent reality, and that the proposition is made true by the fact. (There is a debate in the literature over whether the British idealists really did commit themselves to a coherence theory. See, for example, Candlish (1989).)

As mentioned, alternative (c) sounds like what is otherwise known as the *identity theory of truth*. But there is some reason for caution. The identity theory says that true propositions are identical to facts. But there are two important ways to interpret this idea. Either one can take facts to be primary, and use the theory to reduce true propositions to facts. Or one can take propositions to be primary, and take facts to be no more than true propositions. In the former case, there may be some reason to think that truth is intrinsic to the truth-bearer. But there is an obvious problem concerning how to deal with falsity. False propositions cannot be facts. Hence, an identity theory of the first kind is committed to an oddly disjunctive treatment of true and false propositions. suggestion. The second suggestion faces no problem concerning the existence of false propositions, but it is no longer clear that the identity theory in any way suggests the idea that truth is intrinsic to the truth-bearers. The identity theory thus conceived is fully compatible with the idea that the proposition that p is true (the fact that p obtains) because of how things stand in propositional reality.

If we reject alternatives (b) and (c), the only competitor to (a) is (d). And (d) might, especially when (a) is divorced from other commitments of the correspondence theory, sound unmotivated. If we can simply say that the proposition that *a* is F is made true by *a*'s being F, why accept the defeatist suggestion (d)? But here is one possible consideration in favor of (d). One may think that underlying the attractiveness of (a) is the idea that the proposition that p is true depends on it being the case that p; but it is *necessary* that the proposition that p is true iff p, and one might generally be skeptical of the idea of a dependence relation obtaining between p and q when p and q of necessity obtain together. Given such skepticism—a kind of skepticism of the hyperintensional—there is then reason to favor (d) over (a).

Trenton Merricks (2007) holds that while claims of the form *that p is true because p* are all true, the dependence expressed by these claims is "trivial." and truth does not depend on the world in any "substantive" way. He considers and rejects some more supposedly substantive claims. One is *Truthmaker*, the claim that for every true claim there is some entity whose existence makes that claim true. Merricks rejects this, appealing, for example, to negative existentials. Another is the claim that truth supervenes on being. This thesis receives several precisifications. Merricks argues that under one precisification it is problematic in the way Truthmaker is alleged to be and under another precisification that it fails to spell out the dependence claim. Do Merricks' arguments pose problems for (a)? This is unclear; and it is unclear whether Merricks would even intend them to. Why should the so-called trivial dependence not be sufficient for (a) to be true? Alternatively, if one does find the so-called trivial dependence troublingly trivial, one can contend that Merricks' menu of supposedly more substantive dependence claims is disappointingly short. The point is often made that mere supervenience claims are not strong enough to capture dependence claims (McLaughlin and Bennett (2005): section 3.5). As for Truthmaker, why even contemplate that dependence upon nonpropositional reality should amount to dependence upon the existence of some particular entity or entities in nonpropositional reality?

Alternative (d), the rejection of the question of what makes a proposition true, is a deflationary suggestion, one naturally associated with deflationism. But while this view will be recognized as something a deflationist may hold concerning truth-making, it is, importantly, *not* the canonical characterization of deflationism. The canonical characterization is rather that the content of the notion of truth is somehow *exhausted* by the *equivalence schema*,

(ES) The proposition that p is true iff p

or its counterpart for sentences, the *disquotation schema*,

(DS) "p" is true iff p,

or some more sophisticated variant of one of these schemata. (Sometimes "minimalism" is used for the proposition-based variant of deflationism and "disquotationalism" for the sentence-based one.) Let me next focus on (ES), still assuming that propositions are the primary truth-bearers. In the next section I will discuss what deflationism might come to.

Deflationism

What might it be for the instances of (ES) to exhaust the *concept* of truth? One straightforward thing it might mean is that these instances jointly, implicitly define 'true'. But with the deflationist's characterization of truth thus understood, it is not obvious that deflationism should be understood to entail a deflationary view on the question of the relationality of truth. Why shouldn't the implicit definition have as a consequence that 'true' stands for a property such that whether a proposition has this property depends on nonpropositional reality? There is the illusion, engendered by some of the rhetoric surrounding the deflationist's characterizations of truth, that if the concept of truth is characterized as indicated, it follows that the property of being true is not a full-blooded property. But it is generally not the case that if a predicate is introduced by the stipulation that it makes such-and-such sentences true or such-and-such schemata valid, the predicate does not stand for a (full-blooded) property. Instead, the natural conclusion is simply that the predicate stands for a property which satisfies the conditions mentioned.

Generally, deflationism tends to be characterized in a number of different ways. Sometimes it is characterized by reference to specific types of characterizations of the truth predicate, such as those based around (ES) or (DS). Sometimes it is characterized metaphysically, as it were, as the claim that truth isn't a property, or is only a 'logical' property, or does not have an underlying nature. Sometimes—compare the above discussion of the relationality of truth—it is characterized by its rejecting as misconceived the metaphysical questions to which other traditional theories of truth seek to provide answers. Sometimes in discussions of the philosophy of language it is characterized by its denial of the claim that truth plays any sort of theoretical or explanatory role. It should be clear that the relations between these claims can be problematized.

Suppose that a would-be deflationist insists that when she says that (ES) exhausts the notion of truth she does not mean that it exhausts the *concept of* truth, but rather that it exhausts the *property* of truth. What might this amount to? Surely there will be other facts about the property, not entailed by the instances of (ES) alone, such as that the

proposition that snow is white is true. So there is a sense in which the instances of (ES) clearly do not fully exhaust the property. What else might the claim amount to?

One suggestion is that the view is best understood as the claim that truth satisfies a principle of *revelation* (Damnjanovic 2010; Wright 2001: 753). The property of truth is, according to the claim, fully revealed to us when we grasp the concept of truth. In this it contrasts with *water*. The property of being water is not fully revealed to us when we grasp the concept of water: Further investigation is needed to figure out the nature of this property, namely H_2O. But what does the appeal to revelation amount to? Suppose first that the claim is that the property of truth is such that there is a concept of that property which when grasped fully reveals the property. There is reason to think that this cannot be an adequate explication. Is not every property like that? Isn't it the case that for every property there is *some* possible concept—whether one in our actual repertoire or not—competence with which fully reveals the property? Compare: water has an underlying nature; and the concept water is a paradigmatic example of a concept which doesn't fully reveal the property it stands for, the property of being H_2O. But for all that, there can be some concept which fully reveals this property—indeed, the concept H_2O may be such a concept.

Suppose that instead we take the revelation claim to be that truth is the kind of property such that for *every* concept of a property of this kind, a grasp of this concept fully reveals this property. But isn't this claim implausibly strong? Grasp of the concept of truth may fully reveal the property this concept stands for. But even so, what might be the reason for thinking that this applies to *every* concept of this property?

Someone focusing on the revelation claim can instead urge that the view is that our *ordinary* concept of truth fully reveals the property it stands for. But this claim doesn't by itself yield a distinctive view on the property of truth, if for all properties there are thus revelatory concepts. All we are told is that we happen to have a revelatory concept of this property, while we lack revelatory concepts of other properties.

One way to deal with the problems encountered is to identify deflationism as the view that the only *essential* feature the property of being true is that the instances of (ES) are all true, where 'essential' does not mean simply 'metaphysically necessary'. (Fine (1994) forcefully presents the case for a notion of essence that cuts more finely than the notion of metaphysical necessity.) The concept of truth, understood as the deflationist understands it, is then fully revelatory of the property in the sense that grasp of the concept provides knowledge of all essential features of the property. But talk of revelation is then not needed to characterize deflationism: The claim about the essential features of the property suffices. Notice, however, an irony. Deflationism is often supposed to make truth metaphysically innocuous and safe. But understood as proposed, deflationism carries ideological commitment to a metaphysics of essence.

Truth and the Philosophy of Language

Thus far, I have assumed propositions to be truth-bearers and I have been concerned with a certain metaphysical question: On what does the truth of a proposition depend? I will now turn to the debate over truth within philosophy of language. Any discussion of this must take as its point of departure Alfred Tarski's (1936) theory of truth. For present purposes, the most important features of a Tarskian account are the following: Start with a given language. Let axioms governing simple subsentential expressions state what such expressions refer to, are satisfied by, etc. Then the truth of sentences can

be characterized recursively, using these axioms together with combinatorial axioms. The canonical theorems of Tarski's theory are of the form *s is true iff p*, where s is a name of an object language sentence and p is a sentence of the metalanguage translating s. Tarski proved that such a theory can be consistent, despite the semantic paradoxes. A key idea is that the metalanguage (the language in which the theory is stated) is distinct from the object language (the language which the theory is a theory of).

A central debate concerns the exact philosophical significance of Tarski's theory. An obvious limitation of the theory as it stands is that reference and satisfaction are characterized only by lists: It is not said what reference and satisfaction are; all that is provided are lists that tell us, for a given language, what primitive expressions of that language refer to and are satisfied by. Hartry Field—who early and late has made seminal contributions to the debate—argued (1972) that for Tarski's theory to provide a naturalistically acceptable definition of truth, a suitably reductive general account of reference and satisfaction must be provided, for example along the lines of a causal-explanatory theory of reference. Field also held that thus elaborated, Tarski's theory would be a correspondence theory, albeit one that dispenses with propositions and facts. Another—deflationist—suggestion is to take the list-like characterizations of reference and satisfaction to provide as much information about reference and satisfaction as there is: All there is to reference and satisfaction is captured by these lists. This deflationism would, the idea is, make reference and satisfaction naturalistically acceptable, not through supplying a reductive account but through denying that they possess an underlying nature.

Donald Davidson (1967) made the influential but controversial suggestion that a theory of meaning for a natural language should take the form of a truth theory. The basic idea—which subsequently has been modified in important ways—was that what a speaker who understands a language knows is the information stated by a truth theory. One question concerning Davidson's proposal is whether this is correct. Another question is whether the truth predicate employed in a Davidsonian theory is deflationary or not. The question of whether the truth predicate employed is deflationary in turn splits into two. One question is whether the truth predicate (or the reference and satisfaction predicates in terms of which it is defined) is best conceived of as characterized in the deflationist's list-like way; a distinct question is whether the truth predicate has only a purely expressive role and does not serve a deeper theoretical purpose. While these two questions for natural reasons tend to get run together, they are separate (Eklund 2010).

It is often suggested that there are two fundamentally different outlooks on questions of meaning and content: On one view, truth and related representational notions are to be accorded a central role; on another, such notions have at most an expressive role, and meaning is instead explained by appeal to *use* (Field 1994). There certainly is an important dividing line there. But the picture presented is crude. For one thing, one may suspect that the opposition between alethic accounts and use-theoretic accounts is illusory: that the accounts provide answers to different questions. A use-theoretic account can provide the answer to the question of what determines meaning, while a truth-theoretic account is correct in that what is thereby determined is something best understood in terms of truth-conditions. For another, there are many outlooks on meaning—for example dynamic semantics, and different versions of inferentialism—such that it is not straightforward how best to classify them in these terms. On dynamic semantics, the focus is shifted from the propositions expressed by individual sentences to a sentence's

potential to change an information state. But alethic notions are still important, for the information states are still characterized by what they represent. Inferentialism is sometimes presented as an alternative to a truth-theoretic semantics. But truth can still play an important role in an inferentialist story: for example, because what distinguishes inferences from other associations of thoughts is concern with truth.

Sometimes in discussions about the theoretical role of truth, much of the focus is on reference, as in Field (1972). But some theorists take the theoretical role of truth to be importantly different in kind. First, while Davidson early and late has emphasized the role of a Tarski-style truth theory, he has deemphasized the role of reference. He has taken what might be called an instrumentalist approach to the axioms governing subsentential expressions: any such axioms which yield acceptable conclusions regarding the truth of whole sentences are acceptable. Unsurprisingly, given this outlook, Davidson has been quite cavalier about the possibility of radical referential indeterminacy (Davidson 1979). Truth is central; reference is not. Second, in some parts of the literature, truth is perceived as important because it is a norm—a norm constitutively governing the speech act of assertion, or the mental state of believing (Dummett 1981; Velleman 2000; Wedgwood 2002). One reason it is relevant to distinguish between different reasons for seeing truth as a central notion is that this matters for some debates over whether truth is properly perceived as central. If the theoretical justification for centrally employing a notion of truth relies on the idea that reference is an explanatory notion, arguments that reference is radically indeterminate can be damaging; but if the role of truth is conceived of differently, such indeterminacy arguments are less relevant (Eklund 2010).

The relation between these foundational issues in philosophy of language and the debate over the relationality of truth is not straightforward. For example, it should be obvious that one can insist on the correspondence claim regarding the relationality of truth while eschewing the idea that truth has an important role to play in an account of language. Conversely, one can reject as misguided the question of what the truth of a proposition depends on while still basing one's account of language around a notion of truth.

One reason why the debates as I have described them are not tightly related is that the correspondence theory as introduced concerns the truth of propositions while in the debate over the foundations of semantics the issue concerns truth as applied to sentences. Some theorists—see Field (1986) and McGee and McLaughlin (2000)—understand the difference between the correspondence theorist and the deflationist over sentence truth to concern the modal status of instances of (DS): for the correspondence theorist these instances are contingent while for the deflationist they are necessary. The idea that the instances of (DS) are necessarily true does not seem true for ordinary language: surely it seems that 'white' could have meant green, and if it had meant 'green', then 'snow is white' would not have been true, but snow would still have been white. There are two ways for the deflationist, as Field and McGee & McLaughlin conceive of her, to deal with this. One is to operate not with (DS) as stated but with the rather more complex formulation

(DS*) 'p' is true in the language *I am actually speaking* iff p. (McGee 1993: 92)

But an obvious drawback with appeal to (DS*) is that the instances of (DS*) would appear to be necessary on *any* acceptable view, due to the presence of the actuality

operator. Another—the strategy of Field (1994)—is to insist that the only truth predicate we *need* is a deflationary one, such that the instances of (DS) are necessary.

Related to this modal distinction is the idea that for the correspondence theorist there is something we do with sentences that explains the truth of sentences while the deflationist denies this (McGee 2004). For a correspondence theorist focusing on the relation between propositions and reality, questions about what, if anything, explains the semantic features of sentences are not relevant.

Truth and the Liar

Alongside the discussion in metaphysics about the nature of truth and the discussion in the philosophy of language about the relation between truth and meaning, there is the technical discussion about truth, centered on the liar paradox and other semantic paradoxes. I will not attempt to summarize that debate. Let me just briefly indicate the relation between some issues that come up in the literature on the paradoxes and some themes from the discussion of truth in metaphysics and philosophy of language.

The *liar paradox* seems to show that the schemata (ES) and (DS) are not valid. A liar sentence, L, saying of itself that it is not true (e.g. "this very sentence is not true") is such that $L \leftrightarrow \neg Tr(L)$. But then if the relevant instance of (DS), $L \leftrightarrow Tr(L)$, also holds, we have a contradiction on our hands, $Tr(L) \leftrightarrow \neg Tr(L)$. Corresponding reasoning can be carried out for propositions, but in the technical literature the focus is often on sentences.

This apparent invalidity of (ES) and (DS) presents prima facie problems for a number of different views on truth. Of course one way to deal with this threat posed by the liar is to argue that the arguments for the invalidity of (ES) and (DS) are unsound, contrary to appearances. For example, some might deny that there really is a liar proposition; they would say that a liar sentence fails to express a proposition. But if that fails, what might then be said? Let me go through a couple of views threatened by the liar and discuss what routes lie open.

The paradigmatic deflationist takes (ES) or (DS) to somehow fully characterize truth. Her view is then obviously threatened by the liar. Perhaps the *simplest* way for the deflationist to deal with the liar paradox is to retreat from the idea that (DS) or (ES) in a fully general form is definitional of truth, and instead appeal to a restricted version of one of those schemata. But what restricted version might be appropriate? (See Horwich (1998) and, for criticism, McGee (1992).) Alternatively, she might say that (DS) or (ES) is definitional of truth even if some of the instances are untrue. This general type of view divides into different subtypes. First, the deflationist can embrace *dialetheism*, the view that there are true contradictions. She can then say that even if some instances of (DS) and (ES) are untrue, still all such instances are (also) true (Priest 2006). Second, adopting a *meaning-inconsistency view*, she can say that she is primarily concerned with the concept of truth, and say that her view is that (DS) or (ES) somehow gives the content of this concept, even though it leads to paradox (Eklund 2002, 2007). Third, she can say that (DS) or (ES) is somehow definitional of the concept, but properly understood this doesn't lead to paradox but only to the claim that paradoxical sentences like the liar sentences cannot stably be assigned a truth-value (Gupta 2005).

Turn next to truth-theoretic approaches to meaning, specifically of the Davidsonian form, where the canonical theorems are of the form "s is true iff p," as above. Again

the instance for L presents a problem. Again the dialetheist view and the meaning-inconsistency view can be appealed to. For example, some theorists in the Davidsonian tradition take a meaning theory to be that which competent speakers of a language cognize, and several such theories have entertained the idea that what a competent speaker cognizes is an untrue theory, for example due to liar sentences (Patterson 2007; Badici and Ludwig 2007). A different strategy to which appeal has been made, that of Davidson himself (1967: 28f), is to distinguish the truth predicate in the metalanguage that the meaning theory is stated in from the truth predicate of the object language, and insist that all instances of "s is true iff p," where the truth predicate is the metalanguage predicate and p is a translation of the object language sentence, are true. It seems less plausible to defend a truth-theoretic approach by restrictions on (DS): Surely our competence with the pathological sentences does not differ in kind from our competence with nonpathological sentences.

Does the liar paradox pose problems for a correspondence theory of truth? From some presentations of what the correspondence theory involves one might certainly get that impression. Some friends of the correspondence theory take Aristotle's remark, that to say of what is so that it is so or of what is not so that it is not so, is to speak the truth to characterize a correspondence conception, and Aristotle's remark can be thought to receive expression in (ES). But then if (ES) isn't valid, the correspondence conception is doomed. But to say with (a) above that proposition-truth depends on nonpropositional reality is only to say that the truth of the proposition that p depends on is something non-propositional. However natural it is further to say that it depends on whether p, that further claim is an added commitment. Second, more importantly, Aristotle's remark doesn't immediately entail the validity of (ES). It entails at most that 'S' and 'S is true' are always true together and that 'not-S' and 'S is false' are always true together. It can then allow for exceptions to (ES) when 'S' is neither true nor false. So the correspondence theory is not immediately bound up with (ES). This is not to say, however, that the liar paradox does not pose any problems for the correspondence theory. Take the notion of *groundedness*. A sentence which doesn't assert anything about the truth-values of other sentences is grounded, and one which does assert something about truth-values of other sentences is grounded if the process of tracing through what sentences its truth-value depends on terminates at grounded sentences (Kripke 1975: 693f). It is natural to think that if the truth or falsity of a proposition depends on non-propositional reality, then only grounded sentences can be true or false. But intuitively there are counterexamples to this latter claim. Sentences like $\sim(Tr(L)\&\sim Tr(L))$ and $Tr(L)\vee\sim Tr(L)$ are ungrounded but, intuitively, true. There are three possible reactions to this. One is to take the argument at face value, as refuting the correspondence idea as here spelled out. A second is to hold on to this correspondence idea and reject the intuitive judgments about these propositions (Maudlin 2004). A third is to insist that the truth of some ungrounded propositions is after all fully consistent with the correspondence idea. For instance, one can perhaps hold that a sentence saying of itself that it is not both true and untre is made true by the *general* feature of non-propositional reality that nothing can both be and fail to be the case.

A different argument against the correspondence view departs from the so-called *no-no paradox* (Sorensen 2001). Take two sentences that each says of the other that it is not true. The only consistent ways of assigning truth-values to these sentences involves assigning them different truth-values, so, the argument goes, they must have different truth-values. But since they are perfectly symmetric, there is nothing in

reality that determines one rather than the other to be true; so, given their difference in truth-value, we have a counterexample to correspondence.

References

Badici, E. & K. Ludwig (2007) "The Concept of Truth and the Semantics of the Truth Predicate," *Inquiry* 50: 622–38.

Bradley, F. H. (1909) "On Truth and Coherence," *Mind* 18: 329–42.

Candlish, S. (1989) "The Truth about F. H. Bradley," *Mind* 98: 331–48.

Damnjanovic, N. (2010) "New Wave Deflationism," in Pedersen & Wright (2010).

Davidson, D. (1967) "Truth and Meaning," *Synthèse* 17: 304–23.

—— (1979) "Inscrutability of Reference," *Southwestern Journal of Philosophy* 10: 7–19.

Dummett, M. (1981) *Frege: Philosophy of Language*, Cambridge, MA: Harvard University Press, 2nd edn.

Eklund, M. (2002) "Inconsistent Languages," *Philosophy and Phenomenological Research* 64: 251–75.

—— (2007) "Meaning-Constitutivity," *Inquiry* 50: 559–74.

—— (2010) "Rejectionism about Truth," in Pedersen & Wright (2010).

Field, H. (1972) "Tarski's Theory of Truth," *Journal of Philosophy* 69: 347–75.

—— (1986) "The Deflationary Conception of Truth," in G. MacDonald & C. Wright (eds.) *Fact, Science and Morality*, Oxford: Blackwell.

—— (1994) "Deflationist Views of Meaning and Content," *Mind* 103: 249–84.

Fine, K. (1994) "Essence and Modality," *Philosophical Perspectives* 8: 1–16.

Gupta, A. (2005) "Do the Paradoxes Pose a Special Problem for Deflationism?" in J.C. Beall & B. Armour-Garb (eds.) *Deflationism and Paradox*, Oxford: Oxford University Press.

Horwich, P. (1998) *Truth*, Oxford: Clarendon Press, 2nd edn.

Joachim, H. H. (1906) *The Nature of Truth*, Oxford: Clarendon Press.

Maudlin, T. (2004) *Truth and Paradox: Solving the Riddles*, Oxford: Oxford University Press.

McGee, V. (1992) "Maximal Consistent Sets of Instances of Tarski's Schema (T)," *Journal of Philosophical Logic* 21: 235–41.

—— (1993) "A Semantic Conception of Truth?" *Philosophical Topics* 21:83–111.

—— (2004) "Ramsey and the Correspondence Theory," in V. Halbach & L. Horsten (eds.) *Principles of Truth*, Frankfurt: Ontos Verlag.

McGee, V. & B. McLaughlin (2000) "The Lesson of the Many," *Philosophical Topics* 28: 129–51.

McLaughlin, B. & K. Bennett (2005) "Supervenience," in E. Zalta (ed.) *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/supervenience/.

Merricks, T. (2007) *Truth and Ontology*. Oxford: Clarendon Press.

Patterson, D. (2007) "Understanding the Liar," in J.C .Beall (ed.) *Revenge of the Liar*, Oxford: Oxford University Press.

Pedersen, N. & C. Wright (2010) *New Waves in Truth*, Palgrave Macmillan.

Priest, G. (2006) *In Contradiction: A Study of the Transconsistent*, Oxford: Oxford University Press, expanded edition.

Russell, B. (1910) "The Monistic Theory of Truth," in *Philosophical Essays*, London and New York: Longmans, Green and Co.

Sorensen, R. (2001) *Vagueness and Contradiction*, Oxford: Clarendon Press.

Tarski, A. (1936) "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica* 1: 261–405.

Velleman, D. (2000) "On the Aim of Belief," in *The Possibility of Practical Reason*, Oxford: Oxford University Press.

Walker, R. (1989) *The Coherence Theory of Truth*, London and New York: Routledge.

Wedgwood, R. (2002) "The Aim of Belief," *Philosophical Perspectives* 16: 267–97.

Wright, C. (1999) "Truth: A Traditional Debate Reviewed," *Canadian Journal of Philosophy* suppl. vol. 24: 31–74.

—— (2001) "Minimalism, Deflationism, Pragmatism, Pluralism," in M. Lynch (ed.) *The Nature of Truth*, Cambridge, MA: MIT Press.

—— (2003) *Saving the Differences: Essays on Themes from 'Truth and Objectivity'*, Cambridge, MA: Harvard University Press.

2.3
PROPOSITIONS

Scott Soames

Propositions are traditionally taken to be entities that satisfy A1–A3.

A1. Some things are asserted, believed, and known. These attitudes relate those who assert, believe, or know something to that which they assert believe, or know.
A2. The things asserted, believed, and known are bearers of truth and falsity.
A3. Propositions—the things satisfying A1 and A2—are expressed by sentences. The proposition expressed by S can be designated by the expressions ⌜the proposition that S⌝, ⌜the statement/claim/assertion/belief that S⌝, or simply ⌜that S⌝.

Since different sentences may be used to assert the same thing, or express the same belief, and different beliefs or assertions may result from accepting, or assertively uttering, the same sentence, propositions are not identical with sentences that express them. Intuitively, they are what different sentences, or utterances, that "say the same thing," or express the same belief, have in common. The metaphysical challenge posed by propositions is to identify entities fitting this picture that can play the roles demanded of them in our theories.

One such theory is semantics. In semantics, propositions are needed as contents of (certain) sentences, objects of attitudes, entities quantified over ('Bill questioned several propositions Mary asserted') and referents of certain names ('Logicism', 'Church's Thesis', 'Goldbach's Conjecture') plus demonstratives ('That's true'). Propositions are also needed to state the goals of semantics, which include specifying the contribution of semantic contents to what is asserted and believed by speakers. In theories of mind propositions are contents of perceptual and cognitive states that represent the world as being certain ways, and so have truth conditions. As with sentences, the states and their propositional contents aren't identical. Thus, our question becomes 'What are these things, propositions, that different, but representationally equivalent, sentences and cognitive states have in common?'

One natural, though ultimately unsatisfactory, answer is that what they have in common is simply their truth conditions. On this view, propositions are not things that have truth conditions; they are the truth conditions that sentences and cognitive states have. The main problem with this idea is that sameness of truth conditions is not sufficient for sameness of assertion, belief, or semantic content. Semantic theories that pair a sentence S with the set of metaphysically possible world-states at which S is true don't provide enough information to allow one who knows the theories to understand what S

means, or to identify the beliefs expressed, or assertions made, by uses of S. This negative result persists even if we relax the requirement that the world-states—thought of as maximal properties attributed to the universe—be restricted to those that genuinely could have been instantiated, and allow, in addition, metaphysically impossible world-states that can coherently be conceived to be instantiated, and can't be known a priori not to be. This result is generalized in Soames (1987, 2008b) to all theories satisfying (i)–(iii), no matter what they choose as circumstances at which sentences are evaluated for truth.

(i) Semantic contents of some singular terms, including indexicals and variables, are their referents relative to contexts and assignments.
(ii) ⌜P & Q⌝ is true at a circumstance E iff both conjuncts are true at E.
(iii) ⌜∃x Fx⌝ is true at E iff Fx is true of some object o at E

Since (i) is well motivated and (ii)–(iii) are central to theories of truth at a circumstance, we can't identify the propositions expressed by sentences with sets of circumstances at which they are true, no matter how fine-grained the circumstances. In short, although propositions are indispensable, they can't be extracted from truth-conditional theories. Rather, they are what the truth conditions of sentences are derived from. Thus, we need an independent theory of propositions.

The Realist Theories of Frege and the Early Russell

The realist theories of Gottlob Frege (1892a, 1918) and the early Bertrand Russell (1903) provide a starting point. According to both, propositions are meanings of sentences, bearers of truth, and objects of the attitudes. Since they are meanings, Frege and Russell took them to be structurally complex entities the constituents of which are meanings of the constituents of the sentences that express them. This assumption, though natural, led to what they found to be an unfathomable mystery. Just as sentences aren't collections of unrelated expressions but rather have a structural unity that distinguishes them from mere lists, and is responsible for their representational character, so propositions aren't collections of unrelated expression meanings, but have a unity that distinguishes them from mere aggregates of their parts, and allows them to represent the world. This unity is what Frege and Russell found mysterious.

Here is how the problem is put in Russell (1903: 49–50).

> Consider, for example, the proposition "A differs from B." The constituents of this proposition, if we analyse it, appear to be only A, difference, B. *Yet these constituents, thus placed side by side, do not reconstitute the proposition. The difference which occurs in the proposition actually relates A and B, whereas the difference after analysis is a notion which has no connection with A and B.* [my emphasis] It may be said that we ought, in the analysis, to mention the relations which difference has to A and B, relations which are expressed by *is* and *from* when we say A is different from B. These relations consist in the fact that A is referent and B relatum with respect to difference. But A, referent, difference, relatum, B, is still merely a list of terms, not a proposition. *A proposition, in fact, is essentially a unity, and when analysis has destroyed the unity, no enumeration of constituents will restore the proposition. The verb, when used as a verb, embodies the unity of*

> *the proposition, and is thus distinguishable from the verb considered as a term, though I do not know how to give a clear account of the precise nature of the distinction.*
>
> [my emphasis]

Certainly, there is more to the proposition that A differs from B than the fact that its constituents are A, B, and difference. There is also the manner in which the constituents occur. Presumably, this has something to do with the fact that the proposition *predicates* difference of A and B, and so *represents A as being different from B*. Since a list doesn't predicate anything of anything else, it isn't representational.

Is this problematic? Consider sentence (1), in which 'difference' and 'identity' are nouns, 'different' is an adjective that combines with the copula to form the predicate, and 'from difference' is a prepositional phrase modifying the predicate.

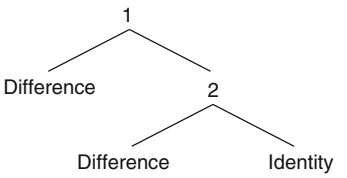
(1) $[_S\ [_N$ Difference$]\ [_{VP}\ [_V$ is different $[_{PP}\ [_P$ from$]\ [_N$ identity$]]]]$

The constituents of the Russellian proposition expressed by (1) are the relations identity and difference, the latter occurring twice. Understanding (1) involves understanding its words, which provide the proposition's constituents, plus understanding its syntactic structure, which indicates what is predicated of what. Just as one who understands the sentence recognizes one expression as predicate and two others as arguments, so, it might be argued, one who entertains the proposition recognizes from its structural configuration which constituent is predicated of which. On this view, it is the *structural relations* that constituents bear to one another that carry the information about predication that *unifies* the proposition, and gives it representational content.

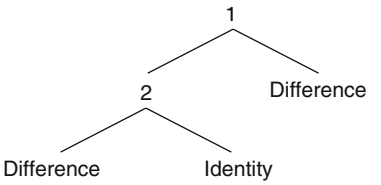
But how, exactly, does the arrangement of its constituents show that the proposition predicates difference of difference and identity? Consider some candidates for this proposition.

(2a) <difference, <difference, identity> >
(b) {{difference}, {difference, {{difference}, {difference, identity}}}}
(c) <<identity, difference>, difference >
(d) {{{identity}, {identity, difference}}, {{{identity}, {identity, difference}}, difference}}

(3a)



(3b)



Although any of these structures could serve as a formal model of the proposition expressed by (1), none could be that proposition. The proposition *represents difference as being different from identity*, because it *predicates* difference of the other two. Since there is nothing in (2), (3), or any similar formal structure, which, *by its very nature*, indicates that anything is predicated of anything, such structures are neither intrinsically representational nor capable of being true or false.

We could, if we wished, *adopt* rules that would allow us to read off information about predication from the structures, and so *interpret* them. But this wouldn't make them traditional propositions. Such propositions *aren't* things we *endow* with meaning by providing them with interpretations; they *are* the meanings, or interpretations, we take sentences to have. The real problem is Russell's pair of assumptions: (i) that propositions are intrinsically representational, independent of us, and (ii) that they are the sources from which cognitive states and sentences inherit their intentionality. To adopt these assumptions is to try to explain something relatively clear and obvious by appealing to something mysterious and incomprehensible. The clear and obvious fact is that agents do predicate properties of things, and when they predicate *being different* of things they—the agents—*represent those things as being different*. The purported explanation is that these agents entertain an abstract structure which—intrinsically, and independent of any relation it bears to us—*predicates* difference of the things, and so represents them that way. The problem with this "explanation" is that we have no idea of how a tree-structure, n-tuple, set-theoretic construction, or any similar abstract structure could, in and of itself, *predicate* anything of anything. Because of this, traditional Russellian propositions are the incomprehensible posits of an explanatory program that doesn't explain. As such, there is no reason to believe in them.

The same holds for Fregean propositions, which are also assumed to be representational independent of us. Frege differs from Russell in postulating "unsaturated" senses that are intrinsically predicative, and so always occur in a predicative role (Frege, 1892b). Although this may sound attractive, it isn't, since it leads him to conclude that neither the sense nor referent of any predicative expression can be designated by a nonpredicative expression—and, thereby, made the subject of a further predication. This thesis—that if Pred is a predicate, then *the sense of Pred* is unsaturated, *the referent of Pred* is incomplete, and neither can be designated by any nonpredicative expression—is self-defeating, as shown by the italicized phrases used to state it (Soames 2010a: chapter 2).

New Conceptions of Propositions

Since we need propositions in our linguistic and cognitive theories, the failure of traditional conceptions calls for a new conception that reverses explanatory priorities. Sentences, utterances, and cognitive states are not representational *because* of their relations to inherently representational propositions. Instead, propositions are representational *because* of their relations to inherently representational cognitive states, or independently representational sentences. On this view, intentionality in mind and language results from the cognitive activities of agents. There are, in the current literature, two main ways of fleshing out this idea. The first, presented in King (2007) takes the existence and representational character of propositions to be dependent on, and derived from, the prior existence and representational character of sentences that express them. The second, presented in Soames (2010a), develops a conception of propositions

as contents of intentional cognitive states generally—including perception and nonlinguistic belief, which are the basis of more complex, linguistically mediated, thought.

King's Conception of Propositions

We begin with sentence (4), and the proposition it expresses.

(4) $[_S\ [_{NP}$ This $]\ [_{Pred}\ [_{Cop}$ is $]\ [_{Adj}$ red$]]]$

The labeled bracketing in (4) displays the syntactic structure of the sentence, indicating its constituents, syntactic structure, and the order in which the constituents occur. Abstracting away from details, I use 'R_S' to name the syntactic relation that the word 'this' stands in to the word 'red' in (4). Since 'this' is an indexical, (4) expresses a proposition only relative to a context of utterance. Imagine a context in which 'this' is used to refer to o. According to King, the proposition—*that o is red*—expressed by (4), relative to the context, is a structurally complex entity of which o and the property *being red* are constituents, along with R_S. He takes this proposition to be the fact designated by (4F).

(4F) the fact that *there is a possible context C and expressions a and b of some language L such that (i) some sentence S of L consists of a's bearing R_S to b, (ii) o is the semantic content of a in C, and the property being red is the semantic content of b in C, and (iii) in L the relation R_S encodes the instantiation function (which is to say that sentences formed by placing an expression α in the relation R_S to an expression β are true, at a context iff the referent of α at the context instantiates the referent of β at the context).*

Although this may seem like a mouthful, the idea is simple. Propositions are what sentences with the same semantic content have in common—the contents of their constituents, plus a common syntactic structure, the semantic contribution of which is the same for each sentence. In English, when a term α stands in R_S to an adjective β, the property designated by β is predicated of the referent of α, which means that the sentence is true iff the referent instantiates the property. So the syntax and semantics of English, plus the existence of (4) and the context C, guarantee the truth of the italicized clause in (4F). King takes this to mean that (4F) designates *a fact*, which is the proposition (4) expresses. Thus, the existence of the proposition that o is red is guaranteed by the semantic properties of (4) at C. To get the same result for any sentence S and context C* at which S has the same semantic properties that (4) has at C, King quantifies over all possible contexts, and every sentence of every language, thereby arriving at the complex general fact that, in his view, is the proposition that o is red.

For King, as for the early Russell, *the fact that a is* G is a complex entity that consists of a's actually instantiating G-hood. As Russell observed, *the proposition that a is* G can't be that fact, for, if it were, its very existence would guarantee its truth (Russell 1912: chapter 12; Soames 2010a: chapter 4). King avoids this absurdity by taking the proposition to be the fact *that a R_p G-hood*—where R_p is the relation expressed by the formula that results from replacing the underlined occurrences of '*o*' and '*the property being red*' in the italicized part of (4F) with a pair of variables. Thus, he solves Russell's problem

of explaining what "holds together" the constituents of the proposition—they are held together by the fact that R_p *actually relates them*—while avoiding the absurd result that in order to exist a proposition must be true.

Although this result is good, the view faces a number of challenges, including (i–iii).

(i) Positing facts—not as true propositions, but as discrete truthmakers—carries philosophical worries that may themselves be sources of skepticism (Merricks 2007; Soames 2008a). Possibilia are also a concern. For King, propositional facts involve quantification over possible contexts, which include merely possible world-states as constituents. *But if the analysis of propositions appeals to a conceptually prior notion of a possible world-state, then the analysis of such a state can't appeal to a conceptually prior notion of a proposition.* This is problematic, since analyses of world-states in terms of propositions are attractive (Adams 1974; Soames 2007, 2010b: chapter 5). Thus, going down King's road on propositions precludes taking other roads one may want to take on related issues.

(ii) King's belief in actualism—which allows reference to, and quantification over, only things that actually exist—brings further worries. He brings possible contexts into his analysis to secure semantic contents (needed for some actual propositions) that no sentence is ever actually used to express. In order for these propositions to exist world-states of possible contexts—thought of as maximal properties the universe fails to instantiate—must also exist. Some of these will, I suspect, involve merely possible objects (including merely possible agents of some contexts). This is worrisome. Though the existence of some uninstantiated properties may be unproblematic, the existence of those involving particular objects (e.g., *being identical with, or distinct from, o*) seems to require the objects to exist. If, as I believe, world-states *are* properties the existence of which requires the existence of possible objects—then their actual existence will require the actual existence of possible objects (Soames 2010b: chapter 5). King takes no definite stand on this, leaving it open how world-states represent possible objects (King 2007: 42–4, 57, 84). This leaves his account hostage to the fortunes of contentious metaphysical claims about which, one might think, the analysis of propositions should be neutral.

(iii) Since English contains both ⌜the fact that S⌝ and ⌜the proposition that S⌝—which for King designate different things—his view seems to require 'that'-clauses to be ambiguous between the readings they bear in ⌜Pam regrets (the fact) that S⌝ and ⌜Pam believes (the proposition) that S⌝. But he neither gives any linguistic argument that this ambiguity exists, nor rebuts seeming evidence to the contrary—for example., "Pam regrets *that she is pregnant*. Although her parents don't realize *it* yet, in time they will come to believe *it*." Here, the fact regretted is described as something that will eventually be believed—a proposition. How, given the supposed difference between the two, can that be?

This difficulty stems from King's Russellian perspective. Both he and Russell take a central aspect of the problem of *the unity of the proposition* to be that of explaining how its constituents—objects and n-place properties—"hold together." Both came to think that the only available explanation is that a relational constituent of the proposition must actually relate them, which means that the proposition must be a fact in order to exist at all. When Russell came to this view, he concluded that the only candidate for being identical with *the proposition* that Desdemona loved Cassio was *the fact* that Desdemona

loved Cassio. Since this ruled out the possibility of falsity, he gave up propositions (Russell 1912: chapter 12). Although King avoids this conclusion by identifying the proposition with a *different* fact, this simply leads to the corresponding difficulty (iii).

The error in both accounts comes from taking the question "What holds the constituents of a proposition together?" too seriously. The misnamed problem of propositional unity isn't that of making one object out of many. Sets, sequences, and trees are each single things with multiple constituents of various sorts. The reason they aren't propositions isn't that their constituents keep falling out. They aren't propositions because they don't represent anything as being any particular way. The real problem for which we have, as yet, no answer is: "How is it that propositions are able to represent the world, and so have truth conditions?" Nothing is gained, when answering this question, by appealing to facts. Even if we suppose that King's clause (4F) designates a complex fact F_4, simply noting F_4's existence provides no hint that *it* has truth conditions at all, let alone that it is true iff o is red. Of course, if it exists, then for some sentence S of some language and some possible context C, S is true at C iff o is red. But from this we can't even conclude that it makes sense to attribute truth conditions to facts, let alone that F_4 has the truth conditions claimed for it.

King eventually realizes this (King 2007: 60). He responds as follows: in and of itself, F_4 isn't a proposition, and doesn't have truth conditions. However, it *becomes* one, and acquires truth conditions, when we *interpret* it in a certain way. King sees F_4 as a complex consisting in *o's* standing in the two-place relation R_p (defined above) to *the property being red*. So understood, it can be seen as acquiring truth conditions, provided that R_p is viewed as (in his words) *"inheriting the semantic significance of* R_S*."* Just as we use the syntactic relation R_S in sentence (4) to *predicate* the property designated by its second argument of the referent of its first argument, so we can *come to use* the propositional relation R_p to *predicate* the property which is its second argument of the object which is its first argument. To do so is to *endow* the formerly nonintentional F_4 with truth conditions. King thinks there was a time—before English contained attitude verbs, 'that'-clauses, or modal operators—when it did contain sentences like (4), plus truth-functional and quantified sentences. *Then*, sentences were used in cognition and communication, words had semantic contents, and sentences had truth conditions. But no propositions yet existed. Although the fact F_4 existed, *it wasn't yet a proposition*, because no need had arisen *to assign semantic significance to* R_p. When the need to report cognitive states was felt, some way of viewing sentences and cognitive states as sharing representational contents had to be found. Then, speakers became aware of F_4, and implicitly assigned it a significance inherited from sentence (4). Thus it came to pass that Man created propositions (King 2007: 60–1, 65–7).

In addition to being uncomfortably speculative, this account embodies a troubling internal tension. On one hand, we are told that propositions are dependent on the *prior* existence of language, and the complex cognitive and communicative practices of those who speak it. Prior to the existence of propositions, speakers are supposed to have used language to think and communicate—which presumably involved (a) using certain expressions with the intention of referring to specific objects, (b) using other expressions with the intention of predicating them of those objects, (c) using sentences to assert things, and express beliefs, and (d) being confident one's hearers could recognize one's intentions, assertions, and beliefs. On the other hand, we are told that belief, assertion, and intention are attitudes to propositions, which did not then exist. From this it follows that the agents who used language in King's prepropositional age must have done so *without intending, believing, or asserting anything*. How can that be?

King answers by postulating that our ancestors had what he calls "proto-intentional states," sufficient for their "prepropositional language." But whatever these "proto-intentional states" are, they can't be relations to things—call them "proto-propositions"—that are themselves representational, or have truth conditions, lest they raise the same problems as genuine propositions do. If the postulated primitive states are not relational in this way, we need to be told: (i) how, if at all, they are representational; (ii) how—if they are not—they can give rise to sentences that are representational, and have truth conditions; and (iii) how—if they both represent things as being certain ways and bring it about that the sentences of the proto-language do too—they can fail to provide propositions at the same time. King's discussion (65–7) doesn't begin to answer these questions. Thus, it remains mysterious how the sentences that, on his view, must have arisen from them could be representational, or have the truth conditions on which his account depends.

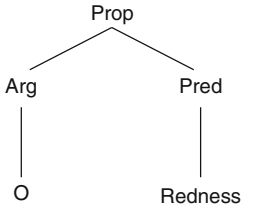
This point is tied to a final question the relevance of which should have been evident all along. What is it to entertain one of his linguistically-based propositions? Apart from points already mentioned, plus an inconclusive paragraph on page 52, King says very little about this. Essentially, all we are told is that once propositions come into being, a speaker of a language that contains attitude constructions will be able to entertain propositions by understanding sentences. No language-independent way of entertaining propositions is specified. But surely, we need an account of propositions as contents of perceptual experience, perceptually based beliefs, plus imaginative, and other non-linguistic, intentional states. It is hard to see how this can come from an analysis of propositions as *inherently linguistic* facts.

The Cognitive Realist Account of Propositions

This observation leads to a second attempt to ground the intentionality of propositions in the cognitive states of agents (Soames 2010a). Like King's approach, it applies to propositions expressed by sentences. Unlike his approach, it is not centered in language but applies to cognitive states generally. We begin with the idea that to entertain a simple proposition is to predicate something of something. To entertain the proposition that o is red is to predicate redness of o. Although, like negation, predication is a primitive notion, it is easily illustrated. When we see o *as red*, we predicate redness of it, and so entertain the proposition that o is red. We also predicate redness of o, and hence entertain this proposition when we form the perceptual *belief* that o is red. We do the same when we understand an utterance of 'This is red', taking the predicate to express the property redness and the subject to refer to o.

Pretend, for a moment, that the proposition that o is red is the abstract structure (4D).

(4D) $[_{\text{Prop}}\ [_{\text{Arg}}\ \text{o}]\ [_{\text{Pred}}\ \text{Redness}]]$



Recognizing structures like these to be theorists' creations, we may temporarily *stipulate* that to entertain (4D) is to predicate redness of o. In so doing, we assign a technical meaning to the verb 'entertain' that explains what we mean by the claim that an agent entertains this abstract structure. Next, we advance (5a), from which we derive that an agent entertains the proposition that o is red iff the agent predicates redness of o.

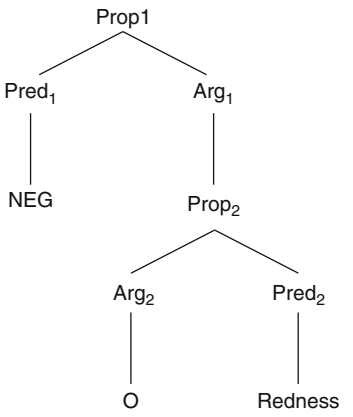
(5a) An agent entertains the proposition that o is red iff the agent entertains (4D).

Other attitudes are treated similarly. Since to *entertain* the proposition that o is red is to predicate redness of o, and since this predication is included in every attitude with that content, entertaining the proposition is one component of any attitude we bear to it. To *judge* that o is red is to predicate redness of o while endorsing that predication. To *believe that o is red* is, (very roughly) to be disposed to judge that it is. To *know* that o is red is, roughly, for o to be red, to believe that o is red, and to be justified in so believing. To *assert* that o is red is to commit oneself, by uttering something, to treat the proposition that o is red as something one knows. Given these characterizations, we add (5b) to (5a), generating further empirical claims that bottom out, as before, in claims about the cognitive acts of agents.

(5b) An agent judges/believes/knows/asserts the proposition that o is red iff the agent judges/believes/knows/asserts (4D).

More complex propositions are treated similarly. For example, to entertain the proposition that it is not the case that o is red, represented by (6), is (i), to predicate redness of o, and thereby to entertain the proposition that o is red, and (ii), to predicate *not being true* of that proposition—which, for the moment, we continue to identify with (4D). One performs this predication—of untruth to the propositional structure—by, in effect, saying to oneself, "That's not true," referring to the result of one's initial predication.

(6) $[_{Prop}[_{Pred}$ NEG$][_{Arg}$ $[_{Prop}$ $[_{Arg}$ o$]$ $[_{Pred}$ Redness$]]]]$



Entertaining more complex propositions involves more complex sequences of cognitive acts.

The approach thus far depends on (i) the inclusion of information about what is predicated of what in the abstract structures we are temporarily allowing to play the role of propositions; (ii) the identification of what it is to entertain a proposition with specific acts of predication that occur in perception and both linguistic and nonlinguistic thought; and (iii) the establishment of propositions as *themselves* representational bearers of truth conditions, in virtue of what is required to entertain them. What makes the proposition that o is red represent o as red, is that predicating redness of o is necessary and sufficient to entertain it. Since one who performs this predication *oneself* represents o as red, and since any agent who does this thereby entertains the proposition, we speak *derivatively* of the proposition predicating the property of the object, and so representing it as red—whether or not anyone, in fact, ever entertains it. Since this fact about the proposition doesn't change from world-state to world-state, its truth conditions are invariant. It is true at any world-state w iff o is red at w.

At this point we confront the convenient pretense that identifies propositions with abstract formal structures like (4D). Having come this far, it is tempting to think that no further, more realistic, account of propositions is needed. The temptation springs from the idea that the function of propositions in our theories is nothing more than to identify and track the cognitive states of agents—which are all we are really interested in. Just as in physical theory we use numbers, and other abstract objects, to specify relations that physical magnitudes bear to one another, so, in semantic and cognitive theory we use propositions to talk about the relations that intentional cognitive states bear to one another, and to the world. Since what is essential is just that we have a simple and economical way of doing this, it is tempting to imagine that one abstract conception of propositions that gets the job done is as good as another. On this instrumentalist view, there are no genuine questions: "What are propositions really?", or "What structures do agents really have before their minds?" Propositions are theoretical fictions.

Although this idea may initially seem attractive, it isn't. According to the theory, entertaining a compound structure like (6)—which is supposed to represent the proposition that it is not the case that o is red—requires agents first to refer to, and then predicate untruth of, its simple propositional constituent, which is the result of their predicating redness of o. But if agents really do refer to the proposition that o is red, and predicate untruth of it, then that proposition is no fiction. Since the representational character of (4D) is due to theorists, and so is merely conventional, neither it nor any similar structure is a proposition. This suggests a genuinely *cognitive realist* view. Since propositions are needed to track cognitive acts, they can be identified with *event types*, instances of which involve those very acts. For example, the proposition *that o is red* is the (minimal) event type in which an agent *predicates redness of o*. As such, it is both intrinsically connected to the cognitive acts it is needed to track, and also something to which all agents who entertain it bear the same natural relation.

Consider a spoken utterance of the sentence 'Snow is white', thought of as an event that occurs at a particular time and place, and also as a token of the sentence uttered. So construed, sentences are event-types capable of having multiple occurrences, which are their tokens. Next imagine an utterance of the sentence followed by an utterance of "That's true." In such a case, the demonstrative may refer either to the utterance, or to the sentence uttered—illustrating that some event types can be bearers of truth value. Finally, there are events in which one doesn't utter anything but simply thinks of snow as white, thereby predicating whiteness of it. These cognitions are events that occur at particular times and places, which are instances of a corresponding event type in which

an agent predicates whiteness of snow. Just as the sentence 'Snow is white' can be identified with an event type of which utterances of it are instances, so the proposition that snow is white can be identified with an event type of which particular acts of predicating whiteness of snow are instances. Thus, both event types have truth conditions.

In addition to bearing their truth conditions intrinsically, propositions-as-event-types are things with which we are acquainted. Since the proposition that o is red is an event type in which one predicates redness of o, and since every attitude one bears to this proposition involves this predication, any agent acquainted with his own cognitive processes—in the sense of being able to make them objects of thought—will be similarly acquainted with the proposition that o is red, by virtue of being acquainted with (and noting the similarity among) the events in his cognitive life that are instances of it. Given the means both of thinking of o as red, and of becoming aware of so doing, one can then make further predications about the content of one's thought. If, after one predicates redness of o, one says to oneself, "That's not true," one thereby predicates untruth of the proposition that is the type of cognitive event one has just experienced. This illustrates how agents are able to entertain compound propositions by predicating properties of their constituent propositions—which was the bane of the instrumentalist view.

In this way, the cognitive-realist theory inherits the virtues of that view, without its deficiencies. Like that view, it provides entities needed as contents of sentences, bearers of truth, and objects of the attitudes. But while the instrumentalist view sees nothing beyond the unavoidably arbitrary formal structures—like (4D) and (6)—that *play the role of propositions*, the realist account views these structures as merely useful devices that represent *the real propositions to which agents bear natural cognitive relations*. The labeled trees provided by linguistic and cognitive theories encode the structure and sequence of cognitive acts that are necessary and sufficient for entertaining the real propositions these structures represent—where entertaining a proposition is performing the acts needed to token the event-type that it is.

This account addresses the most vexing problems to which traditional propositions give rise. Unlike the Platonic epistemology traditionally required by theories of propositions, the cognitive-realist account takes knowledge of propositions to be knowledge of events that make up one's cognitive life. It also avoids the pseudo-problem of "the unity of the proposition," which—though usually posed as that of explaining how the constituents of propositions "hold together"—serves to mask the real problem of explaining how propositions can be representational, and so have truth conditions. The traditional view makes this problem insoluble by taking the representational nature of propositions to be intrinsic, unanalyzable, and independent of us. By locating the representational character of propositions in their intrinsic connection to inherently intentional cognitive events, the cognitive-realist account offers a solution.

References

Adams, Robert (1974) "Theories of Actuality," *Noûs* 8: 211–31.

Frege, Gottlob (1892a), "Uber Sinn and Bedeutung" *Zeitschrift fur Philosophie und Philosophische Kritik* 100: 25–50; translated by Peter Geach as "On Sense and Reference," in Geach and Black (1970).

—— (1892b) "On Concept and Object," *Vierteljahrsschrift fur wissenschaftliche Philosophie* 16: 192–205; translated by Peter Geach, in Geach and Black (1970).

—— (1918), "Der Gedanke," *Beitrage zur Philosophie des deutschen Idealismus* 1; translated by A. Quinton and M. Quinton as "The Thought," *Mind*, 1956, 65: 289–311.

Geach, Peter and Black, Max (1970) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell.
King, Jeffrey C. (2007) *The Nature and Structure of Content*, Oxford: Oxford University Press.
Merricks, Trenton (2007) *Truth and Ontology*, New York: Oxford University Press.
Russell, Bertrand (1903) *Principles of Mathematics*, New York: Norton.
—— (1912) *The Problems of Philosophy*, New York: Oxford University Press..
Soames, Scott (1987) "Direct Reference, Propositional Attitudes, and Semantic Content," *Philosophical Topics* 15: 44–87; reprinted in Soames (2009).
—— (2007) "Actually," in Mark Kalderon, ed., *Proceedings of the Aristotelian Society*, Supplementary Volume 81: 251–277; reprinted in Soames (2009).
—— (2008a) "Truthmakers?" *Philosophical Books* 49, 317–27.
—— (2008b) "Why Propositions Can't be Sets of Truth-Supporting Circumstances," *Journal of Philosophical Logic* 37: 267–76, reprinted in Soames (2009).
—— (2009) *Philosophical Essays*, Vol. 2, Princeton: Princeton University Press.
—— (2010a) *What is Meaning?*, Princeton: Princeton University Press.
—— (2010b) *Philosophy of Language*, Princeton: Princeton University Press.

2.4

CONCEPTS

Christopher Peacocke

Concepts are of interest in their own right; but a good theory of concepts also bears upon a huge array of further philosophical issues. A good theory of concepts has consequences for our understanding of justification, of reasons, the a priori, knowledge, psychological explanation, and the way the mind relates to the world, to name but a few such issues.

'Concept' is used very loosely in everyday English, so we need to be more precise about our target topic. Very roughly, concepts are here understood as meeting the condition that an arbitrary concept C, of whatever category, is distinct from a concept D if and only if a rational thinker can judge a content A(C) without judging A(D) (cp. Frege 1993). So the concept *I* and the concept *Descartes* are distinct, even as used by Descartes. During a bout of amnesia, Descartes can rationally judge *Descartes is Descartes* without judging *I am Descartes*. Similarly the concepts *circle* and *shape described by the equation* $x^2+y^2=r^2$ are distinct, since someone can judge that a seen shape is a circle without knowing what shape that equation describes. Concepts meeting the given condition are to be distinguished from conceptions, stereotypes, and from whatever subpersonal mental representations may underlie or make possible attitudes to contents built up from concepts. For instance, the stereotype of a C.E.O. of a company may be that of a sharply dressed middle-aged man. Nevertheless, a thinker can rationally judge, and may know, that a casually dressed younger woman is a C.E.O of a company.

Just as a theory of meaning and a theory of understanding must be integrated with each other, so equally a theory of concepts and a theory of the grasp of concepts must be integrated with one another. If we give an account of what a concept is, as something that can be true or false of objects, we must also say what it is for concepts so characterized to be grasped, to be in the contents of thinkers' attitudes. Conversely, if we say what it is for a concept to be grasped, we must also say what it is for concepts so described to be capable of being true of false of objects. We cannot gain theoretical understanding if we take the capacity of concepts to play both these roles as something primitive and unexplained (Wilson 2006: Ch 3). Saying that we need a theory that integrates concepts with grasp of concepts is not yet to say which of the two, if either, is more fundamental than the other. That has been a disputed issue between those who would agree that integration is desirable (Peacocke 1992, Fodor 1998).

Dummett draws a distinction between 'meaning theories' which specify the meanings of the expressions in a particular language, and the theory of meaning in general. The theory of meaning which will specify the general form to be instantiated by a particular meaning theory, presumably together with some rationale for that general form (Dummett 1991: 22). I similarly distinguish between theories of particular concepts,

and a general theory which specifies the general form that should be instantiated by a theory of any particular concept. As in the case of linguistic meaning, both types of theory are of philosophical interest.

Theories of intermediate levels of generality, concerned with particular types of concept, can also be philosophically significant. One may be interested in what is common to all observational concepts, or all concepts of conscious states, or all logical concepts.

I have been thinking about a theory of concepts for well over a quarter-century, and could not offer a wholly neutral critical survey even if I tried. So I will not refrain from advocacy; but I will try to indicate, for each alternative approach, some of the tasks it faces. It will help us if we divide theories of concepts into three broad types: type I theories, which hold that there is some constitutive and individuative connection between concepts on the one hand, and reference and truth on the other; Type II theories, which hold that there is no such connection; and Type III theories, whose general character can be indicated by calling them 'no-theory' theories.

Type I Theories

Concepts look very much like Fregean senses. In the *Grundgesetze*, Frege emerges clearly as a Type I theorist (Frege 1967: §32). He says that the Thought expressed by a sentence of his formal language is the thought that the conditions for it to denote the True are fulfilled, where those conditions are determined in a certain canonical fashion by his semantic rules for the formal language. Since Thoughts are composed of component senses, this goes hand-in-hand with the view that particular senses are individuated by the condition for something to be their reference or semantic value, their reference-rules. This conception can be generalized beyond the formal language Frege was considering to a vast range of expressions of natural language.

Simply generalizing this aspect of Frege's conception leaves a lot open, however. All we have so far in this description of concepts is a specification of what a concept is, and not an account of what it is to grasp it. Type I theories that agree on this Fregean component subdivide and part company on the right way of addressing the issue of grasp.

Some theories aim to give substantive accounts of what it is to grasp a concept with a given reference-rule. They hold that there is more to be said than is given in a simple disquotational clause, such as "is in pain" is true of *x* at *t* iff *x* is in pain at *t*. Under the approach I myself favor, for each predicative concept, there a nondisquotational condition tacitly known to one who grasps the concept, and which specifies the condition for an object to fall under the concept. Grasp of the concept *pain* consists in tacit knowledge that an arbitrary subject *x* is in pain just in case *x* is in the same subjective state he himself can recognize when he himself is in pain (Peacocke 2008: Ch 5). Grasp of the observational concept *oval* consists in knowledge that an arbitrary thing *x* is oval just in case *x* is of the same shape as things he himself can recognize in perception as oval (where the representational content of perception captured by *as oval* is nonconceptual). These richer accounts are in the spirit of what Dummett called a full-blooded account, though Dummett's elaboration of that conception contained problematic elements that need not be present in richness (Dummett 1993). (We can proceed without his problematic idea that one could use a full-blooded description to explain a concept to someone who does not already grasp it.) Richer theories of this sort meet the constraint noted earlier. The condition that is tacitly known for something to fall under

the concept is also a priori necessary and sufficient for something really to fall under the concept. Grasp and satisfaction are thereby integrated.

Rejection of such rich theories does not mean that a theorist has nothing to say about grasp. Under the 'modest' conception of McDowell, we may say that someone grasps a concept if attribution of that concept, with its disquotational rule, is required to make the person maximally intelligible to us (McDowell 1998). The issue over richness versus modesty is not whether this condition on maximal intelligibility is correct. It will rather be over whether the constraints on intelligibility can be elucidated without appealing to the resources of a rich theory. Certain combinations of experiential state, environment, and background attitudes will make certain attitudes not intelligibly ascribable to a subject. There are indefinitely many such constraints. The rich theorist will say that these constraints have a finite and structured source in the non-disquotational conditions involved in grasp of the concepts in question. The fundamental issue of whether a theory of concepts should be rich is by no means resolved.

Even if one has sympathy for the rich theorist, the latter has more work to do on several counts. The rich theorist must elucidate the notion of tacit knowledge on which he is relying. He must say how it functions in explanation, both of a thinker's making particular judgments, and in our assessment of certain combinations of attitude and circumstance as intelligible or as unintelligible. He must make it plausible that there really are rich accounts for each of the concepts we employ. He must also say something about the entirely general notions of reference and truth on which he is relying.

Friends of possible worlds semantics will remark that many aspects of concepts are captured if, following Chalmers, they are treated as epistemic intensions (Chalmers 2002). This approach uses a notion of an epistemically possible world. The epistemic intension of an expression is a function from such worlds to extensions. It is the function that maps any such world *w* to the extension the expression would have if *w* were actual. Such an approach has also to explain grasp of a concept. The advocate of this approach could say that grasp of the concept consists in tacit knowledge of the principle associating such a world with the extension the concept would have if the world were actual.

One challenge this approach faces is that of saying why the worlds are needed at all. The treatment must involve some principle determining, for each expression, an extension from a world (if it is actual). Whatever that principle is, it seems that it already fixes a concept or sense, without the need to mention worlds. With such a modification, we obtain a version of the Fregean view that concepts are individuated by their fundamental reference rules. All the same, the approach via epistemic intensions is valuable in explaining the consequences of the Fregean approach when modal operators are in play.

Some accounts of concepts that do invoke reference and truth do so by specifying referential constraints that apply on top of a level that itself does not immediately involve reference and truth. Before we can consider such mixed approaches, we have first to consider treatments of concepts that contain only that reference-free level.

Type II Theories

The most common theories of concepts that aim to eschew reference to reference are conceptual role theories. These theories hold in common that a concept is individuated by its role in a thinker's psychology. This approach treats individuation of a concept and the account of its grasp simultaneously. Within this very general category, there are multiple varieties of conceptual role theory. Some aim to individuate a concept by what

a thinker treats as evidence for certain contents containing it. Others, more pragmatist in spirit, mention the consequences in a speaker's thought and action of accepting certain contents containing the concept.

Cutting across this distinction are two others. Some theories are intrinsically normative. They aim to individuate concepts in terms of good reasons for moving to, or moving from, acceptance of certain contents containing the concept. The other distinction that cuts across all of the above is the line between those theories ('individualist' or 'internalist') that attempt to individuate a concept solely in terms of states which, it is claimed, depend only on conditions within the skin of the thinker, as opposed to those anti-individualist approaches that take into account environmental and possibly social relations of the thinker who employs the concept. Influential conceptual role theories falling within one or another of these categories have been advocated by Sellars (1974), Field (1977), Block (1986), Brandom (1994), Harman (1999), and many others. Prinz (2002) offers a theory inspired by classical empiricism, in treating concepts as a certain kind of perceptually derived representation.

Conceptual role theories have proved tempting in part because it is highly plausible that there are some constitutive links between the nature of particular concepts and some aspects of their use by someone who possesses them. We ought to recognize some special link between observational concepts and their application on the basis of perception; between concepts of sensation and a thinker's own experience of the sensations in question; between particular logical and arithmetical concepts, and absolutely minimal competence in the inferential and numerical application of these concepts.

Acknowledgment of these points does not, however, by itself support conceptual role theories of concepts. For two questions are left unanswered by these acknowledged links. Are the connections with some part of conceptual role consequential on something not itself explained in terms of conceptual role? And, are the connections with conceptual role exhaustive of what individuates the concept?

On the first question, some links with conceptual role are equally explained on reference-involving theories. Suppose, as above, that a thinker's grasp of the observational concept *oval* is given by his tacit knowledge that an arbitrary thing *x* is oval just in case *x* is of the same shape as things he himself can recognize in perception as oval. This tacit knowledge will be manifested precisely in a special sensitivity in his application of the concept to things he perceives to be oval. Similarly, if a thinker has tacit knowledge that any content of the form A&B is true just in case A is true and B is true, that will be manifested in his basic inferential practices involving the concept of conjunction. Similar points can be made for the aspects of conceptual role intrinsic to other concepts.

On the second question, about the exhaustiveness of conceptual role, it is clear that at least the particular features of conceptual role cited here fail to individuate fully the concepts in question. We know what it is for objects too large or too small for us to perceive to be oval. We know what it is for subjects other than ourselves to be in pain. It is very implausible that in these cases, there is a conceptual role, not making reference to reference, that captures these aspects of the concept in question. It is a wholly empirical matter what would be evidence for a proposition that some object too small to be perceived by us is in fact oval. It also a wholly empirical matter what would be the consequences of its being oval. A Quinean point applies here: what such evidence or consequences might be depends on other empirical information. What the evidence or consequences are is not something simply written into the identity of the concept *oval*.

The same applies to concepts of sensations when applied to others; to many past-tense contents; and to many other contents that we judge.

On Type I approaches to concepts, these phenomena are entirely to be expected. We appreciate that for an object we cannot perceive to be oval is for it to be the same shape as the perceptibly oval things. For another subject to be in pain is for her to be in the same subjective state you are in when you are in pain. There are complex and interesting preconditions for grasping such conditions, such as having a conception of objective space, and of multiple subjects of experience. These need philosophical elucidation; but if that can be done, there is nothing illegitimate about such explications of understanding and grasp of the concepts involved. These explications do however place us in the territory of Type I theories. They make essential use of the notion of the property picked out, either in local or in first-person cases, and they make no sense without some such commitment. The attractiveness of such explications suggests that reference has a deeper role to play in the individuation of concepts than is often recognized.

We can distinguish various grades of involvement of reference in an account of concepts. In the conceptual role semantics of Sellars and Brandom, it plays no ineliminable role. At an intermediate level, we can conceive of mixed theories that claim to individuate concepts by conceptual role but which also require these roles to meet some general constraints. One such constraint might be that the transitions specified in the individuation of the concept be at least truth preserving as an a priori matter. The mixed accounts may also try to connect the determination of reference to this truth-preservation. Such was the approach I investigated in the late 1980s and early 1990s (Peacocke 1992). At a third level, of more far-reaching involvement of reference with concepts, as illustrated in the treatments of the preceding paragraph, we cannot even specify what makes something the concept it is without bringing in referential relations in explaining what determines its extension. Correspondingly, we cannot elucidate grasp without mentioning tacit knowledge of conditions that involve reference.

Type III Theories

Type III theories hold that it is a mistake about the nature of concepts to seek a substantive theory of concepts, a mistaken goal pursued both by proponents of theories of Type I and by proponents of theories of Type II. The most developed form of 'no-theory theory' of this third type is that offered by Schiffer (2003). Under an approach of this style, the basic way in which we introduce talk of concepts is by regarding 'John falls under the concept *is in pain*' as a way of rewriting 'John is in pain'. According to Schiffer, this is a something-from-nothing transformation. While it is legitimate to talk of an ontology of concepts, atomic concepts cannot on his view be used to explain our grasp of complex contents, nor can they be used in explanation of our understanding of language. On Schiffer's view, the meaning of 'John is in pain' is given in advance of, and cannot be explained by, anything having to do with concepts. The idea of an explanatory theory of concepts is largely a mistake, according to his Type III theory.

Under this Type III treatment, it is illegitimate to explain a person's understanding of a language compositionally, by associating concepts with atomic expressions. In his exchanges with me, Schiffer has defended this aspect of his view, and the reader is referred to that literature (Peacocke 1986, Schiffer 1986). Here I focus on another challenge to such Type III theories. When we possess a particular concept, *is in pain* or any other, we are able to assess, of an indefinite variety of circumstances, whether those

circumstances make it rational to apply the concept to an object given in a particular way, or whether they make it rational to apply the negation of the concept to the object so given, or whether they make it rational to remain neutral on the issue. We have an open-ended capacity to make these assessments. Since our minds are finite, there must be some finitary body of information on which we draw, no doubt subpersonally, in making these assessments of rationality.

Type I theories holding that possession of a concept consists in tacit knowledge of a substantive, nondisquotational reference can provide precisely such a finitary explanation. The thinker draws on the information given in that substantive, non-disquotational reference rule, and combines that with his other information and capacities that go beyond what is constitutive of the concept to reach an assessment of the rationality of judging the content in the specified circumstances. This gives an answer to the question that any theory should, in one way or another, answer: 'how does the concept-possessor reach that assessment?'

Schiffer's is not the only conceivable theory of Type III. One can imagine a variety of minimalism about meaning which, by analogy with minimalist theories of truth, takes the schema "The sentence 'A' means that A" as primitive, and tries to build an account of meaning from that initial resource. Such a treatment would not take complex meanings as substantive composed from meanings of component expressions, just as minimalists about truth do not regard instances of the disquotational schema as genuinely explained by facts about the references of component expressions. All such views face challenges in explaining language understanding, and our ability to assess the rationality of attitudes, of the same kind as face Schiffer's treatment.

Concepts and Rule-Following

We have a certain intuitive conception of applying a concept in a new case. In the new case, the application of the concept is made for reasons, and the reasons are good (or bad) in the light of the nature of the concept being applied, together with the facts of the new case. This intuitive conception of rational application of concepts is challenged by the sceptical arguments expounded by Kripke (1982) and Wright (1978). These arguments imply that we cannot cite anything nontrivial that makes one rather than another judgment involving a given concept correct in a new case. If these skeptical arguments are sound, then the whole idea of the correctness of a judgment in a new case being a result of the facts of the new case, together with the identity of the concept being applied, has to be rejected.

Skeptical treatments commonly involve a communal or anthropocentric replacement for the intuitive conception of correctness and rationality in a new case. Such skeptical solutions abandon a certain form of objectivity of judgment, in that they hold that the correctness of a judgment is not constituted independently of one or another person's willingness to make the judgment.

Under the conception of concepts as individuated by their fundamental reference rules, there are straight, nonskeptical answers to these skeptical arguments. A judgment applying a concept in a new case to a given object is correct if the object meets the condition given in the fundamental reference rule for the concept. The judgment is rational if it is reached by the thinker's drawing on her tacit knowledge of the concept's fundamental reference rule. That knowledge constitutes her grasp of the concept. So under this account, it is after all grasp of a concept that contributes to the explanation

of judgments made in new cases. One can expect the explanation to draw also on other information possessed by the person applying the concept. It is only in a restricted class of cases that the explanation of concept application draws only on a thinker's grasp of a concept.

Those who endorse the intuitive conception of rule-following are sometimes accused of commitment to an unacceptable Platonism. What is said to be unacceptable is a conception of correctness that outstrips, even in ordinary cases, anything we could be supposed to know or to access. But it is a mistake to think that a proper rejection of the inaccessible, in the description of ordinary cases, forces us into the rejection of the minimal objectivity of judgment. The understanding-condition for *oval* given above involves the notion of something's being of the same shape as the perceptibly oval things. The holding of such a relation of sameness of shape is something we can on occasion know to obtain; it is not inaccessible; and it is not something judgment-dependent either.

Concepts and the a Priori

Just as Quine gave compelling arguments that nothing is true purely in virtue of meanings, one can equally argue that nothing is true purely in virtue of the concepts composing it (Quine 1976). This point is still consistent with the existence of a deep connection between the nature of concepts and the property of a content of being a priori—that is, the property of being knowable without justificational reliance on perceptual experience or other conscious states.

When some state gives a reason for judging a conceptual content, that reason-giving relation is founded in the nature of the concepts that compose the conceptual content (Peacocke 2000). A genuine perception of something as oval can give reason for judging that the object thus perceived is oval. The fundamental reference rule for the concept *oval* explains why this should be so. It follows from that reference rule that when such a perception occurs, the relevant conceptual content *That's oval* is true.

Here we need to use the notion of the relatively a priori. It is relatively a priori that the content *That's oval* holds, given the occurrence of a genuine corresponding perception of the object in question as oval. Similarly, the occurrence to a subject of a pain justifies him in judging *I'm in pain*. It is relatively a priori that the content *I'm in pain* holds, give the occurrence to the thinker of a pain. Again, the fundamental reference rules for *pain* and the first person concept *I*—that it refers to the agent of the thinking containing it—explain why these justification relations hold.

This is not at all a conventionalist approach. Indeed, in appealing to the nature of concepts, the approach is apparently explaining necessary relations (instances of the relatively a priori) from other certain constitutive truths (the fundamental reference rules for concepts).

The absolutely a priori contents form a limiting case of the relatively a priori contents. They are those contents whose acceptance is justified on the basis of certain reasons, such as proofs, where the justifying reasons do not involve perceptual or other conscious events or states, and where the reasons can be ratified as sound on the basis of the fundamental reference rules for the concepts from which the contents are composed. Logical and mathematical theorems arguably meet this condition. In developing this approach to the absolutely a priori, it is important to respect the distinction between what gives access to a justifying condition, and the justifying condition itself. We perceive proofs of logical and mathematical theorems. The perception gives access to the proof. What

justifies is the abstract proof itself. Though proofs are perceived, the justifications to which these perceptions gives access are not empirical.

It remains a challenge for the theory of concepts to discover how much of the a priori can be explained by such resources.

Concepts and Relation-Based Thought

In some familiar cases, a subject is able to think of an object under a given concept only because he stands in a certain relation to that object. The relation makes available the concept as a way of thinking of the object. This is widely acknowledged for perceptual-demonstrative concepts of particular objects, such as *that man, that tower*. It is also widely acknowledged for recognitional concepts of particular individuals and for perceptible properties and relations. But in fact the phenomenon is far more widespread than cases of these sorts.

For every recognizable kind of conscious psychological event or state, there is a concept of it made available by the thinker's ability to recognize when she is in that state, or enjoying that event. Such concepts plausibly include our concepts of particular types of sensation, perceptual experiences, and action-awarenesses. (Recognizability, which involves memory, is of course less fine-grained than conscious discriminability.) Our everyday concept of seeing, for instance, is not that of a state that plays a specified explanatory role. Many people have false beliefs about the role in explanation played by conscious experience. They think it explains the properties of concurrent action, when in fact properties of such action are explained by nonconscious states sustained by the older, dorsal route to representation (Milner and Goodale 1995). But those false beliefs do not mean that their owners are not really exercising the concept of perceptual experience. They are exercising that concept because of the recognitional relations in which they stand to perceptual experience, not because of any beliefs they may or may not have about the explanatory role of experience.

It is not only spatial, material, and mental states that can be thought about in ways made available by the relations in which a thinker stands. Concepts (and meanings) themselves can be thought about in ways made available by the thinker's grasping those concepts. A thinker can fully grasp the second-level concept *the concept man*, *the concept parliament* only if she already grasps the concepts *man* and *parliament* respectively. Moreover, just as in the example of the concept of perceptual experience, a thinker can possess these concepts of concepts even if she has radically false beliefs about what it is to possess or to be a concept.

What makes a thinker's concept *the concept F* is not any theory she may have about concepts, but rather the relation of this second-level concept to her own grasp of the concept *F* itself. Suppose that *F* has the fundamental reference rule R. In grasping *F*, the thinker has tacit knowledge of the reference rule R. To possess the concept *the concept F* is then to have tacit knowledge that the condition for an arbitrary concept φ to be the reference of *the concept F* is for φ to have the fundamental reference rule R. This tacit knowledge is, in the nature of its content, available only to someone who grasps the fundamental reference rule R. A similar point can be made about meanings.

Since philosophers and linguists disagree frequently about what individuates particular concepts and meanings, it could hardly be the case that correct theories of concepts and meanings are necessary for thinking about a concept *F* as *the concept F*. The generalization of a relation-based treatment to the case of concepts and meanings

explains why here, as elsewhere, false theory is consistent with a uniform notion *the concept F*.

Concepts and the Metaphysics of a Domain

How should we conceive of the relation between a theory of concepts of a given domain (of things, events, and properties) on the one hand, and the metaphysics of that domain on the other? One Dummett-like thesis about the explanatory priority of a theory of concepts generalizes what Dummett says about the theory of meaning. The thesis states that metaphysical disputes in general can be resolved within, and should really be regarded as disputes about, the theory of concepts (Dummett 1991: 13–15). An opposing view holds that the metaphysics of a domain is explanatorily prior to the nature of concepts of that domain. I argue here for this opposing view, which I label 'metaphysics-first'.

The correct metaphysics for a domain constrains accounts of grasp of concepts of objects and properties in that domain. It does so because it constrains the possible relations in which a thinker can stand to elements of that domain, and to the instantiation of its distinctive properties and relations. There is a requirement of consonance of our theory of grasp of concepts with the metaphysics of what falls under those concepts.

The metaphysics of some domains allows for the intelligibility of the causal operation of objects, events, and properties in that domain upon thinkers. The primary class of such cases is the domain of material objects, events, and their properties and relations in space and time. These mind-independent things can stand in perceptual relations to thinkers, and thus make available distinctive, perceptually based concepts of them. Perceptual-demonstratives, recognitional concepts and observational concepts all have fundamental reference rules that depend on and make reference to these perceptual relations.

Some conscious mental events stand in causal relations to their owners, without standing in perceptual relations. A visual experience, or a pain, may rationally cause the engagement of a thinker's recognitional capacity for just that kind of conscious event. Such an account of some mental concepts requires an ontology of events falling under them, rather than a treatment of such events as some kind of construct from expressive action.

Unlike physical and mental events and their properties, other domains, such as those of possible worlds and possible objects, do not stand in causal relations to any mental states of thinkers, or to anything else. A fortiori, the fundamental reference rules for such notions cannot mention causal relations to their subject matter. In such cases, it is much more plausible that the fundamental reference rules must have a more intimate relation to the very metaphysics of the subject matter. One general model for such cases is that the thinker has some tacit knowledge of what makes something a possible world, or a possible object. That knowledge is drawn upon in the thinker's assessment of the truth-value of thoughts whose truth conditions involve reference to possible worlds or possible objects.

In this area, as everywhere else, we need to make use of the distinction between the order of discovery and the order of philosophical explanation. For example, it may be that in the first instance we cannot conceive, epistemically, of two distinct possible objects with all of the same relations to actual objects and to other possible objects. We may first conceive of this as an epistemic impossibility. But on reflection and further philosophical theory, we reach the conclusion that the epistemic impossibility is

founded in a metaphysical impossibility. Possible objects are individuated in such a way that there is no genuine metaphysical possibility of the kind in question. Here the metaphysics remains prior in the order of philosophical explanation, even if it was not temporally prior in the order of our thought.

There remain many outstanding issues about the relation of epistemology to metaphysics, issues for which the theory of concepts is pivotal. Epistemic norms in a given domain are a wide range of cases traceable to the nature of the concepts in the domain. If the nature of those concepts is given by their fundamental reference rules, and those rules are in turn constrained by the metaphysics, it follows that the metaphysics contributes to the explanation of, and constrains, epistemic norms. It remains to be seen how this dependence plays out in detail across a huge range of domains, from such abstract domains of thought about the infinite and the possible on the one hand, and on the other to thought about consciousness, subjects of experience, and norms themselves.

References

Block, N. (1986) "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy* 10, 615–678

Brandom, R. (1994) *Making It Explicit: Reasoning, Representing and Discursive Commitment* (Cambridge, MA: Harvard University Press)

Chalmers, D. (2002) "On Sense and Intension," *Philosophical Perspectives*, 16, 135–182

Dummett, M. (1991) *The Logical Basis of Metaphysics* (Cambridge, MA: Harvard University Press)

Dummett, M. (1993) "What is a Theory of Meaning? (I)" in his collection *The Seas of Language* (Oxford: Oxford University Press)

Field, H. (1977) "Logic, Meaning and Conceptual Role," *Journal of Philosophy* 74, 379–409.

Fodor, J. (1998) *Concepts: Where Cognitive Science Went Wrong* (Oxford: Oxford University Press)

Frege, G. (1967) *The Basic Laws of Arithmetic* tr. and ed. M. Furth (Berkeley and Los Angeles: University of California Press)

Frege, G. (1993) "On Sense and Reference" in A. Moore, ed., *Meaning and Reference* (Oxford: Oxford University Press)

Harman, G. (1999) *Reasoning, Meaning and Mind* (Oxford: Oxford University Press)

Kripke, S. (1982) *Wittgenstein on Rules and Private Language* (Oxford: Blackwell)

McDowell, J. (1998) "In Defence of Modesty", in his collection *Meaning, Knowledge and Reality* (Cambridge, MA: Harvard University Press)

Milner, D. and Goodale, M. (1995) *The Visual Brain in Action* (Oxford: Oxford University Press)

Peacocke, C. (1986) "Explanation in Computational Psychology: Language, Perception and Level 1.5," *Mind and Language* 1, 101–123.

Peacocke, C. (1992) *A Study of Concepts* (Cambridge, MA: MIT Press)

Peacocke, C. (2000) "Explaining the A Priori: The Programme of Moderate Rationalism" in *New Essays on the A Priori* eds., P. Boghossian and C. Peacocke (Oxford: Oxford University Press).

Peacocke, C. (2008) *Truly Understood* (Oxford: Oxford University Press)

Prinz, J. (2002) *Furnishing the Mind: Concepts and their Perceptual Basis* (Cambridge, MA: MIT Press)

Quine, W. (1976) "Truth by Convention" in his *The Ways of Paradox and Other Essays* (Cambridge, MA: Harvard University Press, Second Edition)

Schiffer, S. (1986) "Peacocke on Explanation in Psychology," *Mind and Language* 1, 362–371

Schiffer, S. (2003) *The Things We Mean* (Oxford: Oxford University Press)

Sellars, W. (1974) "Meaning as Functional Classification", *Synthèse* 27, 417–437

Wilson, M. (2006) *Wandering Significance: An Essay on Conceptual Behavior* (Oxford: Oxford University Press)

Wright, C. (1978) *Wittgenstein on the Foundations of Mathematics* (London: Duckworth)

2.5

ANALYTIC TRUTH

Cory Juhl and Eric Loomis

Introduction and Overview

'Analytic truth' is 'truth in virtue of meaning', broadly construed. Statements that are not analytically true are typically regarded as 'synthetic' statements for which truth or falsity is determined, in part, by meaning- or language-independent matters of fact. Other categories of statement, such as analytically false statements, may exist as well. This chapter will briefly survey some of the historical uses of analytic truth, as well as objections to it raised by W. V. O. Quine and others. We then turn to some contemporary discussions of analyticity, presenting our own account of analytic truth and examining other recent accounts.

Historical Conceptions of Analytic Truth

Statements such as:

(i) All bodies are extended;
(ii) Everything is either B or not B;
(iii) If A is earlier than B and B is earlier than C, then A is earlier than C;
(iv) Nothing is both red and green all over, and;
(v) 1 + 1 = 2

have appeared to philosophers at different times to possess any one of a variety of properties, and have all been labeled 'analytic truths'. Immanuel Kant regarded (i) as analytic on the grounds that the concept expressed by its predicate, 'extended', is 'covertly contained' in that expressed by its subject, 'body' (Kant 1965: 48–9). This containment implied, Kant thought, that analytic truths are knowable a priori, since one does not "go beyond the concept 'body'" in order to find extension "bound up with it." Moreover, such analytic truths are necessary, in the sense that the denial of an analytic truth such as (i) would lead to a contradiction because, presumably, it would require saying that extension both was and was not a property of body (1965: 189). For Kant however, a priority and necessity are not only properties of analytic truths but also features of some synthetic ones, such as the truths of arithmetic which he believed to be synthetic a priori (53).

Subsequent philosophers such as Bernhard Bolzano found Kant's theory of analytic truth to be both vague and narrow. Bolzano described Kant's notion of containment

as "figurative" and permitting "too many interpretations" (1973: 201). Kant's notion was too narrow because it excluded statements such as (ii) from being analytic, for there is no apparent sense in which being B or not-B is "contained" in the concept of everything, even "covertly" (201). Bolzano thought a statement like (ii) ought to be counted as analytic, for no matter what changes we make to the constituent non-logical concept 'B', the statement remains true. Bolzano thus introduced a third feature commonly ascribed to one sort of analytic truth—preservation of truth upon the uniform substitution of nonlogical expressions.

However, Bolzano recognized that even this idea is too narrow to capture a statement like (iii) within the extension of analytic truth, since (iii)'s truth makes essential use of the meaning of the nonlogical expression "is earlier than"; if we uniformly replace this expression by certain others, such as "is the mother of," it ceases to be true. Nor does a statement like (iii) appear to be analytic either by Kant's notion of containment or by his claim that a statement is analytic just in case its denial leads to a contradiction. Yet (iii) appeared to later philosophers to possess certain features, such as necessity and knowability a priori, also supposedly possessed by analytic truths. Moreover, in the early twentieth century, the emergence of more powerful systems of logic, due to Frege, Russell, Hilbert, and others, as well as the emergence of axiomatic systems of arithmetic and related formal sciences, raised the possibility of a further expansion of the scope of analytic truth to include such statements.

Many philosophers, most famously those of the Vienna Circle of logical empiricism, thus awakened to the possibility that all necessary a priori statements, including those of arithmetic, could be given a uniform characterization, thereby removing the need for Kant's distinction between analytic and synthetic a priori truths. Vienna Circle members such as Rudolf Carnap, and sympathizers such as A. J. Ayer, expanded the notion of analytic truth to include all statements that are necessary and knowable a priori. Their doing so involved three important modifications to the notion of analytic truth. First, they treated analytic truths as expressions of linguistic rules which constitute the meanings of the expressions of a language. A statement like (iii) could thus be regarded as partially constitutive of the meaning of "earlier than" (cf. Carnap 1937; Ayer 1952). Even axiomatic systems for arithmetic or other formal sciences were seen as rules for the use of the corresponding terms and the concepts thereby expressed. Second, they considered analytic truths to be empty of "cognitive content" and thus "tautologies" in the sense advanced by Wittgenstein in the *Tractatus* (Ayer 1952; Carnap 1956). That is, while analytic truths may reveal aspects of linguistic usage *qua* linguistic rules, they say nothing about how the world is. Third, Vienna Circle members accounted for the a priori status of analytic truths by regarding them as *conventions* which were known a priori in virtue of their status as stipulations (cf. Ayer 1952: 80; Carnap 1967: 178). The Vienna Circle and those influenced by it thus treated analytic truths as possessing a variety of characteristics: true by convention, knowable a priori, expressing rules for the use of words, and "tautologous" or empty of cognitive content.

Objecting to Analytic Truth

In the early 1950s, Willard V. Quine launched a series of criticisms against the notion of analytic truth that were later developed by others such as Gilbert Harman. Behind Quine's criticisms lay his belief that philosophy ought not to be regarded as fundamentally distinct from the natural sciences (cf. Quine 1960: 3–4, 24). For many,

analytic truths appeared to have a distinctive status, namely to be such that their truth is independent of matters of language-independent facts, including empirical facts. Analytic truths thus appeared to reveal a sphere of inquiry that could be conducted independently of empirical science. So by criticizing analytic truth, Quine could be seen as partially undermining any distinctive role or domain of inquiry for philosophy.

Quine's criticisms of analytic truth took a variety of forms. Some of them rested on Quine's view that intensional notions, such as meaning and synonymy, are philosophically suspect. For example, so-called Frege analytic truths, in which a statement is analytic if it can be converted into a logical truth by the uniform substitution of synonymous expressions, rely on intensional notions and therefore deemed by Quine to be illicit (Quine 1953: 21–2). Another line of criticism, later developed by Harman, suggested that analytic truths simply don't exist as a matter of fact; "analytic truth" [is like a] witch" (Harman 1967: 127). According to this objection, there may be agreement among philosophers about which statements are called analytic truths, just as there may have been agreement in some communities about which women were to be called witches. Yet the extensions of both terms may turn out to be empty.

A third line of Quinean objections to analytic truth involved Quine's endorsement of a holistic conception of belief and confirmation that derived from Duhem, Carnap, and others (cf. Carnap 1937: 317; Quine 1953: 43). According to this holism, a statement can have empirical content only within a broader system of theory, a system which may include the entire "fabric" of our beliefs (Quine 1953: 42). An individual hypothesis *H* can only be verified given a theoretical background of statements of theory (such as empirical laws), or other observations, methodological assumptions, mathematical truths, and logical laws by which *H* may be inferred from the other statements. If *H* is empirically tested and found to be false, any element of the theoretical background in which it is embedded is, in principle, open to possible rejection. Thus, upon finding that *H* is false, we may reject the supporting observations, or the statements of theory, the methodological assumptions, or even the mathematical or logical laws used in inferring *H*. Among these statements are some that the tradition regarded as analytic truths. Yet we see how they may nonetheless be empirically rejected. As such, no statement is immune to revision in the way that analytic truths were thought to be. Given any alleged analytic truth, circumstances can be imagined in which subsequent empirical discoveries lead to its rejection by leading to modifications of the "web" of belief in which it is embedded (cf. Quine 1953: 43).

In recent decades, another challenge to analytic truth has emerged from the rise of "semantic externalism," deriving from work by Saul Kripke, Hilary Putnam, and others. Kripke (1972) argued that the reference of proper names, like "Gödel," and of natural-kind terms like "gold" or "water," depends upon factors external to any given language user, such as causal and historical factors. Such considerations arguably undercut core elements of the notion of analyticity as traditionally conceived (by the Vienna Circle, for example). First, many philosophers were persuaded by Kripke and Putnam to cease regarding the notions of a priori knowledge and necessary truth as coextensive. In particular, Kripke and Putnam argue that some necessary truths can be known only a posteriori, and that some a priori truths might be contingent. Kripke argued that seemingly a priori truths, such as that the standard meter stick in Paris is one meter long, might turn out to be contingent. For while we might claim to know this a priori on the grounds that it is stipulated, this should not be taken to show that it is necessarily true, Kripke argued (1972: 55f). That stick *S* is one meter long is only a contingent feature; it may

after all have a different length in some counterfactual situations in which it is heated or stretched. If "necessary truth" is not coextensive with "truth knowable a priori," then an analytically true statement cannot be said to imply both necessity and a priori knowability.

A second consequence of semantic externalism for analytic truth concerns the idea that an analytic truth is one that is true in virtue of meaning, or one which someone who understands the sentence must be disposed to assent to. This idea loses its relevance if the meaning of a statement is not something that a competent user of a language can be said to know just in virtue of their linguistic competence. Consider again Kripke's gold example: the meaning of "gold" is given through an a posteriori discovery by empirical science, namely that gold has the atomic number 79. This was not known until relatively recently. Nonetheless, English speakers prior to that discovery were competent in their use of the word "gold" (a fact explained on Kripke's account by saying that those speakers nonetheless understood the reference-fixing explanations for "gold"). Semantic externalism thus raises a challenge to the idea that possessing linguistic competence in the use of certain expressions might be sufficient to account for either the knowledge of or the truth of those expressions.

Several philosophers have recently explored these and other consequences of semantic externalism for analyticity. Some have argued that semantic externalism renders the notion of analyticity largely uninteresting (see Salmon 1993; Williamson 2007). Williamson in particular has developed a series of arguments designed to raise doubts concerning the view, originating with the Vienna Circle, that analytic truths are non-cognitive, tautologous, or trivial. Williamson separates the majority of his objections from those of Quine; unlike Quine he accepts synonymy and allied intensional notions as legitimate (2007: 52–3). Yet, like Quine, he finds the notion of analytic truth to be of little philosophical interest. Whether analytic truths are construed "metaphysically" as truths whose meaning is sufficient for their truth (60–1), or "epistemologically" as expressing truths knowable in virtue of meaning (77–8), Williamson argues that the notion of analytic truth achieves few of the explanatory goals required by the philosophical "linguistic turn" deriving from the Vienna Circle and its successors.

A Stipulative Conception of Analyticity

The objections to analytic truth that were raised by Quine, as well as those deriving from semantic externalism, have led many contemporary philosophers to view the notion with skepticism. Yet for a variety of reasons, variants of the notion continue to be defended. Here is one: Let us say that a *statement* is a sentence-as-understood by a community of speakers. A statement stands to a sentence roughly as a chess bishop stands to a mini-sculpture of a bishop. When we describe a mini-sculpture as a chess bishop we are treating it as an item understood to be governed by rules of use associated with a particular activity, the game of chess. Similarly, when we describe a sentence as a statement, we understand it to be governed by rules of use shared by other speakers. For present purposes we focus on the relatively idealized situation in which all speakers of the language understand the statement in the same way, and take the associated sentence to be such that everyone takes it to be governed by the same rules.

A statement is *analytic* in our sense if its associated sentence is understood as governed by rules of the following form:

> (Stip) Sentence *s* expresses a true proposition. Furthermore, that *s* expresses a true proposition is empirically indefeasible.

In addition we introduce a stricter notion of analyticity that we will use when we discuss mathematical and related applications, that of a transcendental stipulation:

> (TStip) Sentence *s* expresses a true proposition. Furthermore, that *s* expresses a true proposition is empirically indefeasible. Finally, the proposition *p* expressed by *s* is also empirically indefeasible.

We introduce the notion of a transcendental stipulation in order to distinguish Kripke-type cases such as the stipulation (J), "Julius is the unique inventor of the zipper (if there is a unique inventor of the zipper)," from stipulations that "express" propositions that are understood as empirically indefeasible. The proposition expressed by (J) is a proposition whose truth or falsehood is supportable by empirical evidence, even though the proposition *that (J) expresses a true proposition* is understood as empirically indefeasible. In contrast, in mathematical and many other cases we take the proposition expressed by the stipulation-sentence s to be empirically indefeasible (for details, cf. Juhl and Loomis 2010: ch. 6).

Responses to Some Objections to Analyticity

As we noted, Quine rejected all "intensional" notions and entities, and so refused to accept accounts of analytic truth that relied on them. However, since most contemporary philosophers accept intensional notions, we will assume them in what follows. Nonetheless, we wish to consider how the account introduced above fares against two popular objections to analytic truth.

> Objection 1: Synonymy is indeterminate (or there is no relation of synonymy). Since analyticity relies on a notion of synonymy, analyticity is similarly ill-defined.

The account of analyticity given above does not require appeal to synonymy. The characterization appeals to the notion of a rule, and assumes as a constraint on acceptable translation schemes that they preserve clear cases of analyticity. Many stipulative definitions are universally quantified biconditionals (UQBs) of the form "For any x, Fx iff Gx." It seems natural to say that F is synonymous with G when everyone agrees on the stipulation. A focus on this sort of example may be what leads to the impression that stipulations in general generate synonymies. Yet there is a more general phenomenon that should be captured, the phenomenon of the adoption of a novel rule for a language, one which concerns the empirical indefeasibility of the associated sentences. UQB stipulations are relatively easy to understand as empirically indefeasible, since these logical forms do not entail "categorical" statements that can then potentially conflict with experience. Nevertheless, we think that the notion of a rule of the form Stip is in the background in both these typical UQB cases and many other cases that need not have this logical form. What is required for a coherent linguistic practice is that the rules of the practice be coherent. In order to maintain coherence, all rules added concerning

stipulations must be such that the collection of stipulations as a whole does not provide evidence either for or against any possible experience.

The stipulative account also appeals to the notion of an acceptable translation scheme. Any good scheme should translate any clearly analytic statement into a clearly analytic statement, understood as governed by rules of that distinctive sort. This constraint insures that no analytic statement will count as given up on the basis of empirical counterevidence. Quineans appeal to principles of good or better translation, even though they avoid appeal to synonymy. However, there is no need for a detour through synonymy in order to defend analyticity.

> Objection 2: Saying or believing that sentence s is true does not make s true. The truth of any sentence is only justifiable on the basis of its playing an essential role in some best/good empirical theory of the world.

Williamson raises a version of this objection using what he calls "understanding-truth links." On one view of stipulations, what it is for stipulations to be true simply is for people to have a convention of taking them to be true. According to such a view, there is no gap, for certain sentences, between our believing that they are true and their being true. Our sharing a belief in their truth, or adopting a convention of taking them to be true, constitutes their being true. This is analogous to other cases such as whether some item is "really" a chess bishop or another is "really" a stop sign. There is arguably nothing more to being a chess bishop or a stop sign beyond our having some collection of intentional states concerning the items. Stipulations and their truth might be thought to be similar. If we permit stipulations in our language, and our practice is coherent, then one could claim that there are trivial arguments demonstrating the truth of relevant types of stipulations. In a mathematical proof, for example, if at some stage the mathematician says, "f(a) = b," and writes as justification "stipulative definition," then given the rules of proof, the truth of the sentence "f(a) = b" might count as proved.

It seems possible to remain agnostic on the question of whether stipulations within a coherent stipulation practice are true, at least while reflecting on the practice, as opposed to engaging in it. However, this agnosticism is compatible with recognizing analytic truths' philosophical interest. We may show why given our conventions it seems natural to describe them as true (whether or not they "really are" true), and why it is not puzzling that we could take them to be true in a distinctively risk-free way, i.e., without risk of empirical counterevidence. The practice is designed to permit such risk-free assertions. Perhaps this is itself a matter to be resolved by stipulations governing the extension of "true."

The second part of this objection states that all sentences are to be justified in ways broadly the same as all empirical statements are. This seems to us to be an optional, and misleading, picture of how our language works, and a narrow way of understanding how any possible language practice *can* be understood.

Mathematics and Other Applications of Analyticity

Thus far we have focused on stipulative definitions. Another possible case for analyticity is mathematics. The current debate about the epistemology of mathematics is dominated by two main positions: Platonist-type realists on the one hand, and empiricists influenced by Quine and Mill on the other. Many philosophers are uncomfortable

with an account that appeals to mathematical intuition, as realism tends to do. Yet mathematical empiricisms do not seem to plausibly capture an obvious feature of mathematical practice: namely that mathematicians ignore empirical evidence of the sort that empirical scientists require for their research. It is not that empiricists about mathematics have not tried. But the accounts given of which we are aware have not proven satisfying for many philosophers of mathematics. Stipulationism (about mathematics—we will often omit the qualifier) offers an alternative to these views. Stipulationists can maintain that mathematical axiom-statements share the epistemic, pragmatic, and semantic profiles of transcendental stipulations, i.e., that they *are* transcendental stipulations.

One can motivate stipulationism by noting that understanding mathematics as "analytic" sheds some light on the epistemology of mathematics and makes some problems seem less pressing. For example, how do we know that there is a set of concreta {a,b} whenever the concreta a and b are known to exist? It seems odd to say that we intuit the existence of such an entity. It is also difficult to imagine setting up an experiment to find out whether the set really exists over and above the members. It seems instead that there is something like a stipulative definition of "set" such that a certain conditional is understood as true or taken to be true in the way that stipulations are, so that given any concreta x and y, there is a set containing exactly them. Stipulationism might be extended to other cases of abstracta such as fictional objects, mereological sums, and propositions (see Schiffer 2003; Juhl 2009). There are three objections to stipulationism about mathematics that we will consider here.

> *The "existential" objection*: You cannot stipulate things into existence. The existence or nonexistence of Santa Claus, God, or other objects is a matter of stipulation-independent fact.

The kinds of objects this objection invokes as examples are the sorts of things on which empirical evidence is understood to bear. Santa's existence is something for or against which we can have empirical evidence; similarly for God, at least as classically conceived (as a performer of miracles, for example). This is relevant because on a stipulationist view of mathematics, the members of the language community understand mathematical statements in such a way that the possibility of empirical evidence (the existence of experiences with some character) bearing on their truth or justification is ruled out. Speakers are also understood to make other statements for which empirical evidence is not ruled out. If a given class of statements were understood to be such that they both permitted experience to bear on them, and simultaneously ruled this possibility out, the resulting practice would be incoherent. Thus, in any coherent stipulation practice, that is, any language that permits the introduction of empirically indefeasible stipulations, stipulations cannot jointly have consequences on which empirical evidence bears.

The upshot of a response to objection 1, then, is that for some kinds of things, or some concepts, certain statements and/or facts involving them are understood as being evidentially independent of experience. For such statements and their logical consequences, a coherent stipulation practice can permit the introduction of statements with arbitrary logical form, even existence claims. In that sense, at least, existence claims can be stipulated. And the standard Santa Claus, unicorn, and God counterexamples do not undermine this point (cf. Juhl and Loomis 2010).

> *The "mathematics uses empirical evidence" objection*: Empirical evidence frequently counts for or against the truth of mathematical statements. For example, the fact that a computer prints a certain pattern of ink, that a mathematician utters a pattern of sound, or that a pattern appears in a textbook can all be empirical evidence for the truth of various mathematical statements.

Williamson and others think that "experiments" are relevant to mathematical truth (2007, p. 6). However, it is worth drawing a distinction between the way in which empirical evidence bears on mathematical statements versus empirical hypotheses. We can distinguish between first-order and second-order justifications. A first-order or "canonical" justification for any (nonaxiomatic) mathematical statement consists of a proof of it. No empirical statement is permitted to be appealed to in a proof. The fact that a computer produces a physical output in the relevant cases is contingently correlated with the existence of a first-order justification, and provides second-order justification. Similarly for an utterance of a mathematician; we only count the utterance as evidence if we think that, say, the fact that this person utters that p is correlated with the existence of a first-order justification. In short, empirical evidence can bear on the truth of a mathematical statement. But unlike ordinary empirical hypotheses, empirical evidence bears only in a particular way: namely in this "second-order" way that provides evidence for the existence of a first-order, purely nonempirical justification.

> *The "indispensability" objection:* Mathematics is as indispensable to physical science as electrons and quarks are, and this fact provides empirical evidence that mathematical entities exist.

This objection sees the support that physical science grants to the existence of objects like electrons as extending to mathematical "entities." However, the fact that the entities referred to by the mathematical-looking terminology applied in physics are distinctively *mathematical*—outside of space and time, knowable independently of experience, and the like—plays no role in the explanations and calculations used in physical science. For consider that we can take any theory of mathematized physics, and in effect "Ramseyfy" it (i.e., "quantify out" all predicate terms with second-order quantifiers, and similarly for constants and objectual quantifiers). The resulting theory would be exactly as empirically adequate as the original theory. Precisely what parts of the Ramseyfied theory would be empirically supported in the general case is a complicated matter. But it is perfectly coherent to say that in such a Ramseyfied theory there is no term referring to anything distinctively mathematical. Thus, even if every part of the Ramseyfied theory is taken to be empirically defeasible, the stipulationist practice in which properly mathematical statements are treated as empirically indefeasible would remain coherent.

The problem that we are suggesting for the indispensability argument, then, is this: the entities and predicates being distinctively *mathematical* does no predictive or explanatory work. Thus, the claim that mathematical entities *per se* are indispensable is not supported by successful empirical applications of mathematical language. All that is genuinely supported is the existence of *some structures or other*.

We will take the Quinean view to argue that one *must* be empiricist about mathematics. Stipulationists hope to show that one *need not* be empiricist about mathematics. Stipulationism yields a description of mathematics that fits better with actual practice of mathematics and actual scientists than the alternative, empiricist stipulation.

This debate about the nature of mathematical statements connects to the general Quinean picture according to which all statements are empirically defeasible. The stipulationist thinks that the case of mathematics exposes Quine's view as an optional proposal, one that does not fit actual mathematical or scientific practice.

Other Recent Developments

Recently Gillian Russell has produced an important book on analyticity (Russell 2008). Russell takes as her paradigm case the sentence "I am here now." There is a clear and natural sense in which token utterances of that type are (almost) always true, and true "in virtue of the meanings of the words therein." Russell takes the notion of analytic truth to be, at bottom, the notion of a statement or sentence whose truth is explainable solely by appeal to something akin to meanings of the component words. Her account of analytic truth is thus one primarily concerned with semantic issues, not epistemological ones, although she grants the possibility that competent speakers may know the truth of an analytic claim solely through knowing facts about "reference determiners" of its component words. As such, Russell's account, unlike the stipulative account above, is compatible with the possibility that a broadly Quinean account of mathematical and other allegedly a priori truths is correct.

According to another approach to analyticity, (cf. Peacocke 1995; Boghossian 1996; Glock 2003), linguistic or conceptual competence with some term or concept requires belief in the truth of some statements. In some recent works, both Boghossian and Peacocke have argued for a priori justification of these meaning-conferring principles on the basis of their special meaning-conferring status. Against such views, Williamson argues that linguistic competence may never require belief in any statement or proposition (2007: 73ff). And even if the competence requirement is correct, a number of critics (Salmon 1993; Horwich 1998; Williamson 2007) have argued that there is no way to show that the statements that are meaning-conferring are therefore true or a priori justified.

One of Williamson's arguments against the view that linguistic competence entails believing certain statements considers two philosophers who do not believe that all vixens are vixens (Williamson 2007: 85ff). One rejects the principle because he thinks that conditionals whose antecedents and consequents are truth valueless are themselves truth valueless. This philosopher thinks that some entities x are such that a statement that x is a vixen is truth valueless. Williamson argues that this case and others can be constructed in such a way that we will judge that the person in question understands the notion *vixen* (as well as implication and universal quantification). Thus linguistic competence does not entail belief, even for this very basic logical truth. Williamson thinks that this strategy can be generalized, showing that no belief is ever absolutely necessary for linguistic or conceptual competence.

Closely related to the claim about linguistic competence is the view that various principles "implicitly define" some terms, and that their role in conferring meaning entails that the defining principles are true. Horwich gives an argument against the related idea that meaning-conferring principles must be true (Horwich 1998: 136ff). Horwich notes that novel theoretical terms can be introduced by "defining" the terms by appeal to theoretical principles or hypotheses that are taken to hold true for them. It is plausible to describe such principles as "conferring meaning" on the terms that they are intended to define. Yet consider the classic case of a now-rejected theoretical type,

phlogiston. The term "phlogiston" had some theoretical principles that were taken to hold of it, and which arguably conferred meaning on the term. Yet as it turns out, there is no phlogiston. To the extent that the theoretical principles entailed that phlogiston was the cause of some event type, the principles are false. This remains true, even if we grant that those very principles somehow "define" or confer meaning upon the term "phlogiston." Since Horwich thinks that the analytic/synthetic distinction was primarily introduced to resolve some epistemic problems associated with a priori knowledge or justification, he takes the phlogiston example to show that there is no interesting analytic/synthetic distinction.

It is possible, Horwich thinks, to grant that the "non-substantive" conditional of the form "If anything has (theoretical profile) T, then phlogiston has T" is knowable a priori, in sharp contrast to the truth of the "substantive" defining statements, which are the consequents of the "trivial" conditional. We think that this tempting move is misguided. First of all, Horwich does not explain what he means by "substantive" when he says that logical principles and mathematical principles are substantive, whereas the conditional above is not. One might think that since the conditional is not logically true, if logical truths are already substantive, as Horwich thinks they are, then such a conditional is even more substantive. In any case, given Horwich's sympathy with Quine's confirmational holism and Quine's rejection of a priori knowledge, it seems that Horwich (and other Quineans) should think that even the conditional might be revised under some imaginable empirical circumstances. The conditional has the form, "If any theoretical concept has feature T, then phlogiston does." Suppose that physicists were to find that, after all, there is some fluid that satisfies T (its presence causes temperature to rise, etc.), but that this fluid is not the fluid that physicists would have been interacting with using their very low-tech instruments of the nineteenth century. The fluid that does satisfy T requires sophisticated twenty-seventh century superstring-detection devices to interact with it systematically. In such a situation, physicists arguably should deny the conditional, "If anything has T, then phlogiston does," since something satisfies T, as it turns out, whereas phlogiston does not exist. The example is not much more or less far-fetched than many Quinean thought experiments purportedly showing that some analytic statement "might be given up." Furthermore, any attempt to rule out scenarios by putting more and more information concerning phlogiston into T makes it less plausible to claim that everything within T is *required* for conferring meaning on the term. Thus we think that according to Horwich's Quinean background assumptions, even the conditional is "revisable," and not a priori justified. His move from the ordinary stipulation to the conditional seems to be a mistake.

That said, the conditional *can* be coherently stipulated. That is, a conditional of this form *can* be held to be true and empirically indefeasible, if it does not have empirical consequences by itself. Like the UQBs we discussed earlier, such a conditional has a logical form that is readily immunized from counterevidence. However, mathematical statements are similarly easily immunized from the empirical. And if it introduces rules for a novel term into the language, the consequent of a Horwich conditional can be stipulated as well, as long as it does not bear on even the probability of the truth of any empirical statement. An important difference between the empirical cases like phlogiston and mathematical cases is that, in practice, scientists actually do give up theoretical principles taken to be true of a theoretical notion, even if that notion was initially introduced or "defined" via its role in an empirical theory. Mathematical practice contrasts sharply with empirical science practice in this respect.

Examples like phlogiston are interesting but often overgeneralized. One can grant that for properly theoretical notions, our practices as a matter of fact seem to be of a sort such that practically any principle originally introduced as "defining" could eventually be rejected on the basis of empirical counterevidence. In contrast, as a matter of fact mathematicians practically never justify the rejection of any defining principles on the basis of empirical evidence. So generalizing from empirical practice to mathematical practice is overgeneralization. Horwich seems to think that there are substantive and nonsubstantive statements, and that mathematics and logic fall on the substantive side. We are instead inclined to draw a distinction between statements understood as empirically indefeasible and statements understood as defeasible, and to place mathematics in the former category and most statements of empirical science in the latter.

None of the criticisms against analyticity to date are sufficient to rule it out in every form, and many of the disputes in ontology and metaontology may benefit from a much closer and more sympathetic treatment of analyticity, such as the recent accounts canvassed above. There remains space for analyticity to have an important and epistemically illuminating role, even if it cannot be deployed as ambitiously as some philosophers have thought.

References

Ayer, Alfred J. (1952) *Language Truth and Logic*. New York: Dover Publications Inc.

Boghossian, Paul. (1996) "Analyticity Reconsidered," *Noûs*, 30:3, 360–391.

Bolzano, Bernard (1973) *Theory of Science*. B. Terrell (translator). Dordrecht: D. Reidel.

Carnap, Rudolf (1937) *The Logical Syntax of Language*. London: Routledge and Kegan Paul.

Carnap, Rudolf (1956) "Empiricism, Semantics, and Ontology." In *Meaning and Necessity, A Study in Semantics and Modal Logic*. Enlarged Edition. Chicago: University of Chicago Press, 205–221.

Carnap, Rudolf (1967) *The Logical Structure of the World and Pseudoproblems in Philosophy*. Second Edition, R. A. George (translator). Berkeley: University of California Press.

Glock, H. J. (2003) *Quine and Davidson on Thought, Language and Reality*. Cambridge: Cambridge University Press.

Harman, G. (1967) "Quine on Meaning and Existence I," *Review of Metaphysics* 21, 124–151.

Horwich, P. (1998) *Meaning*. Oxford: Oxford University Press.

Juhl, C. (2009) "Pure and Impure Stipulata," *Philosophy and Phenomenological Research*, LXXIX: 3, 637–652.

Juhl, C., and Loomis, E. (2010) *Analyticity*. New York: Routledge.

Kant, I. (1965) *Critique of Pure Reason*. N. K. Smith (translator). New York: St. Martin's Press.

Kripke, Saul A. (1972) *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kripke, Saul (1980) *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Peacocke, Christopher (1995) *A Study of Concepts*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1953) "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 21–46.

Quine, W. V. O. (1960) *Word and Object*. Cambridge, MA: MIT Press.

Russell, Gillian. (2008) *Truth in Virtue of Meaning*. Oxford: Oxford University Press.

Salmon, N. (1993) "Analyticity and Apriority," *Philosophical Perspectives* 7, Language and Logic: 125–133.

Schiffer, S. (2003) *The Things We Mean*. Oxford: Oxford University Press.

Williamson, Timothy. (2007) *The Philosophy of Philosophy*. Oxford: Blackwell.

2.6

POSSIBLE WORLDS SEMANTICS

Daniel Nolan

One approach to specifying the meaning of pieces of languages is to treat those meanings as constructions out of possible worlds and possible objects. This technique is useful both in logic and in providing the semantics of natural languages. After introducing possible worlds semantics, this chapter will outline some of the applications that have convinced many philosophers and linguists of the usefulness of this framework, and will conclude with a discussion of one of the most conspicuous limitations of the framework, and some ways this limitation has been dealt with.

Extension and Intension

Pieces of language have extensions: predicates are associated with sets of objects that satisfy the predicate, sentences have truth-values, and singular referring expressions designate objects. Some aspects of the meanings of expressions seem to rely on more than the extensions of those expressions: extensions are often contrasted with "intensions." In the context of possible worlds semantics, expressions are assigned possible worlds intensions that are constructed from possible worlds and possible objects using set theoretic means. For example, the standard possible worlds intension of a predicate is a function from possible worlds to extensions that yields the predicate's extension in each possible world. (These functions are sometimes called the *properties* associated with predicates.) One standard way of defining the intensions of declarative sentences is as a function from possible worlds to truth-values (a function that assigns a sentence True when it is true according to a world and False otherwise), or, as a simpler construction that carries the same information, the set of possible worlds where the sentence is true. These sets of possible worlds are often referred to as possible worlds *propositions* expressed by sentences.

Possible Worlds Semantics in Logic and Language

The expression "possible worlds semantics" was first used to describe "semantics" in the logician's sense. In this sense, possible worlds semantics is a matter of associating with a given logic a *model* that contains worlds; and assignments, relative to those worlds, of the truth-values of sentences, extensions of predicates, and so on. The best-known

application of possible worlds semantics is in the semantics of modal logics, usually attributed to Kripke 1959 and 1963, though see Copeland 2000 for the "prehistory" of this semantics.

Consider, for example, the propositional modal logic S4, which extends propositional logic with two operators applied to well-formed formulas: $\Box(\phi)$ for "it is necessary that ϕ" and $\Diamond(\phi)$ for "it is possible that ϕ." A model of S4 can be given by a ordered triple <W, R, ν>, where W is a set of worlds, R is a two-place relation on worlds, and ν is an evaluation function from propositional letters and worlds to the truth-values 1 and 0 (for True and False, respectively). R is an "accessibility relation" between worlds, and to ensure the model will model S4, we insist that R be both reflexive (so that for all $w \in W$, Rww) and transitive (so that for all w, x, $y \in W$, when Rwx and Rxy then Rwy). To specify ν, we first insist that ν assign exactly one of 1 and 0 to each propositional letter for each $w \in W$. We will represent ν assigning a propositional letter p the truth value 1 at world w as follows: $\nu_w(p) = 1$, and in general assignments of truth-values at worlds will be represented in the form $\nu_\alpha(\phi) = \gamma$ for a world α, proposition ϕ and truth-value γ.

Next, we specify that ν assigns propositional-logical compounds truth-values, given the truth-value assignments to propositional letters in each world, as those truth-values are built up in propositional logic. So, for example, when $\nu_w(p) = 1$ and $\nu_w(q) = 0$, then $\nu_w(p \,\&\, q) = 0$, $\nu_w(p \vee q) = 1$, $\nu_w(\sim p) = 0$, $\nu_w(q \supset p) = 0$, and so on. Finally, we specify how formulas with and ◊ are to be treated by ν. When ϕ is a well-formed formula, $\Box(\phi)$ is to be assigned the truth value 1 at a world w by ν ($\nu_w(\phi) = 1$) provided that all the worlds v such that Rwv are also such that $\nu_v(\phi) = 1$. In other words, a sentence ϕ is specified to be necessarily true at a world w if and only if in every possible world *accessible* from w (or "possible from" w) is a world where ϕ is true. The condition for ◊ is that ($\nu_w(\Diamond\phi) = 1$) provided that *some* world v such that Rwv is also such that $\nu_v(\phi) = 1$. In other words, ϕ is possibly true in a world w just in case some world "possible from" w is one where ϕ is true.

Propositional S4 is provably *sound* and *complete* with respect to these models. Whenever a conclusion Γ is provable from a set of premises Σ in S4, then every model which satisfies the above conditions and in which all the premises Σ are true at a world is one where Γ is also true at that world. Conversely, whenever it is that case that, in every model of the above sort, Γ is true at every world where all of Σ are true, then Γ is provable from Σ in the logic S4. Because of these facts, we can use the model theory to tell us which arguments are valid in S4 and which are not: if we can provide an S4 countermodel of an argument from some premises Σ to a conclusion Γ, then we can tell that Γ will not be provable from Σ in our S4 proof theory, for example.

Such models can be complicated to handle predicate calculus versions of S4, as well as still richer versions of S4. The model given above can also be varied to yield classes of models for other modal logics: in particular, we get quite a range of different modal logics just by putting different constraints on the R relation. Possible worlds models like these, whether for propositional logic or for predicate logic, have enabled modal logicians to see connections between a number of different systems, and in particular seem to have helped logicians to get a better understanding of the behaviour of *iterated* modal operators, such as in the formula $\Diamond\Box p$.

Models of this sort have been used in a number of different areas in logic besides the treatment of "necessarily" and "possibly." Similar semantics have been offered for temporal logics (for example the logics of "until now," "from now on" etc.), deontic logics (the logics of "ought," "permitted," and so on), and some epistemic and doxastic logics

(logics of "it is known that . . ." or ". . . is (epistemically) justified"). Note that while the formal structures are similar, the "worlds" in these models are often better thought of as times (in the temporal case) or epistemic states (in some epistemic logics). There is an extended sense of "possible worlds semantics" where the models for a logic are a possible worlds semantics provided that they are similar, in the right kinds of ways, to the Kripke models for the logics of necessity and possibility. Compare the way that different mathematical systems are known as "geometries" even if they bear little relation to the structure of space, provided they bear certain kinds of similarities to systems like Euclid's geometry.

Other kinds of relations on worlds have been found useful by philosophical logicians besides the accessibility relations of normal modal logics. Many nonclassical Relevant Logics use a ternary relation on worlds to model the implication relations in those logics (see Routley et al. 1983). "Conditional logics" like that of Stalnaker 1968 and Lewis 1973 use a "closeness" relation on worlds to model conditionals, particularly counterfactual conditionals. Interpretations of the language of probability sometimes employ *measures* on sets of worlds: This is one way to apply Kolmogorov's mathematical theory of probability.

One thing that it is worth noting about the models used by logicians to investigate the formal properties of logics is that these models can be very useful even if they have nothing in particular to do with the project of providing the *meanings* of the logical operators and connectives: We could still employ the model theory to help with finding counterexamples to invalid inferences, for example. On the other hand, in linguistics and philosophy of language "semantics" is the study of the meaning of expressions (perhaps particularly that aspect of meaning which is in common across different uses of the same word). A number of linguists and philosophers have found it useful to investigate the semantics of natural languages, such as English, employing a framework of possible worlds. "Possible worlds semantics," in this sense, involves using theories that postulate possible worlds in order either to give the meanings of, for example, English expressions, or at least to illuminate how those meanings work.

The two senses of "possible worlds semantics" are connected. Richard Montague introduced possible worlds into the systematic study of the semantics of natural language, and his approach was to take a simplified fragment of English, and then treat it as logicians had treated their artificial formal languages. Montague offered a model theory for his fragment of English that employed possible worlds, possible objects, and various set-theoretic constructions from them, in order to systematically yield assignments of truth-conditions to sentences: that is, assignments of truth-values to sentences relative to each world of the model. His reason for offering this model theory was to exhibit the way that the meanings of different English words go together to produce truth-conditions of English sentences. His semantics, in the model-theory sense, was the core of his semantics, in the sense of a theory about the meaning of English constructions. Montague's system can be, and has been, applied to many other languages besides English, and Montague grammar remains a flourishing research program in linguistics. Montague (1974) is the best place to find Montague's own contributions.

What Are Possible Worlds?

Possible worlds semantics relies on there being a domain of possible worlds, and usually requires things in those worlds to be the members of the sets associated with predicates

in each world. So we have possible worlds and possible individuals. Does that mean that, in order to use possible worlds semantics, we need to think that there is an infinite range of alternate universes, full of merely possible individuals, including strange individuals like golden mountains and talking donkeys?

Some have argued that this is indeed the best way to understand possible worlds and possible individuals (Lewis 1986). More common is the view that possible worlds are some kind of abstract object. Whatever they are, they need to be able to have sentences or propositions "true according to" them, but there are a number of ways this might happen. A world literally containing a singing cow would have "A cow sings" true according to it in one straightforward way. But if worlds are representations, they might be able to have "A cow sings" true according to them in the same way that a newspaper that contains the sentence "A cow sings" can. And there are other options: Perhaps a world composed of states of affairs that do not obtain could have as a component the state of affairs of a cow singing, for example. Or we could offer a modal analysis of "true according to": Perhaps a claim is true according to a world if, and only if, *were that world actualized*, the claim would be true. Finally, a theorist might resist analysis of the "true according to" relation altogether, and just take it to be an unanalyzed fact that "A cow sings" is true according to one possible world but not according to another.

Some logicians and philosophers of language will be happy to leave the question of what possible worlds are, and what truth according to a world is, to the metaphysicians. But it is worth noting that the answer given to what possible worlds are, and in particular what the relation of truth-according-to-a-world is, may put constraints on what theoretical purposes possible worlds are suitable for. If the modal analysis of "true according to" is relied upon, then it will be much harder to treat the equivalence of "necessarily ϕ" and "at all (accessible) possible worlds, ϕ" as *explaining* necessity: we can use modality to explain what possible worlds are, or possible worlds to explain modality, but using each to explain the other risks circularity.

More worryingly, if we treat possible worlds as representations—as sets of sentences, or sets of propositions, or "maximal" propositions that settle every question—then we risk undermining their explanatory role in a theory of meaning. If we are explaining the meaning of sentences by associating them with sets of possible worlds, then on pain of circularity we should not take possible worlds to be sets of interpreted sentences, for example.

There are many other questions that could be raised about possible worlds, and indeed about possible objects. But for some purposes the metaphysics does not matter. For example, when using possible worlds semantics in the logicians sense, whether the members of W or the members of the domains associated with each $w \in W$ are anything like real possible worlds and the full range of merely possible objects is irrelevant from the technical point of view.

Indeed, even if we are using possible worlds for the project of semantics in the linguistics sense, we may be able to treat them as only formal modeling devices. Whether or not we can do so will depend on what sort of explanations they are meant to provide—if they are invoked only for modeling structural constraints on meanings, it may not matter what the objects in the models are. On the other hand, if we take ourselves to be illuminating the meaning of expressions by providing genuine truth-conditions for sentences, then the "worlds" and constructions from them should arguably have something to do with what can possibly be the case.

Possible Worlds Semantics for Natural Language: A Simple Model

Each kind of linguistic construction is assigned an "intension" or "semantic value" constructed from possible worlds, possible objects, and sets ultimately built up from these. *Sentences* are assigned, as their intensions, sets of possible worlds: these are the worlds at which the sentences are true, so in assigning these sets we are specifying the conditions under which the sentences are true. These possible world truth conditions are identified as *propositions* in this framework. *Predicates* are assigned functions from worlds to sets of possible objects: the set associated with a predicate and a world is the extension of that predicate at that world. These functions from worlds to extensions are identified as *properties* in this framework. For an object to *have* a property, in a world, is for the object to be in the extension of that property at that world.

We can provide intensions for quantifiers: the intension of the universal and existential quantifiers are certain function from worlds to sets of *properties*, in the above sense. The universal quantifier, for example, is assigned, at each world, the set of properties in common to *all* the objects at that world, and the existential quantifier is assigned the set of properties such that *at least one* thing at that world has that property.

The intensions of larger linguistic structures are functions of the intensions of their components, so, for example, we can build up the intension of a sentence from the intensions of its constituents. The rule for a sentence of the form [quantifier];[predicate], for example, is that a world w is in the intension of the sentence if the intension of the quantifier is a function from w to a set of properties that includes the property associated with the predicate. This may become clearer with an example: suppose our sentence is "something is hungry." When we evaluate a world, say w_5, the intension of "something" will associate with w_5 a set of properties (the set of properties had by at least one thing in w_5). Suppose the property that is the intension of "is hungry" is a member of that set. Then w_5 will be one of the members of the proposition associated with "something is hungry." Which is as it should be, since those conditions insure that one of the objects in w_5 is in the extension of the property associated with "is hungry." If one of the objects has the property of being hungry in a world, we want "something is hungry" to count as true there.

If we wanted uniform principles about how different semantic categories went together, and the two kinds of basic sentences we allowed were [quantifier];[predicate] and [name];[predicate] sentences, then it would make sense to give names the same general kind of semantic values as quantifiers. Montague (1973) did this, treating names as also being associated with sets of properties: "Barack Obama," for example, would be associated with a function from worlds to the set of properties that Barack Obama has at those worlds. "Barack Obama is hungry" would then have a world w in its intension just in case the set of properties associated with "Barack Obama" at w included the property associated with "is hungry." Alternatively, we could have different rules for different kinds of sentences: For example, we could allow the semantic value of "Barack Obama" just to be an individual, and the rule for [name];[predicate] sentences to be that a world w is in the intension associated with such a sentence provided that the individual that is the intension of the name is in the extension, at w, of the property associated with the predicate.

Extensional connectives such as "and," "or," and "not" can be assigned intensions: "And," for example, can take two propositions as inputs and yield a proposition as output (it, in effect, it functions as a set-theoretic intersection on sets of worlds: ϕ *and* φ is

true at the set of worlds that is the intersection of the set of worlds where ϕ is true and the one where φ is true). Intensional operators like the sentential adverbs "necessarily" and "possibly" can also be treated as functions from propositions to propositions. The simplest rules are these (where ϕ is a sentence): *necessarily* ϕ has, as its intension, the proposition containing every world just in case the intension of ϕ includes every world, and has a null intension otherwise. *Possibly* ϕ has, as its intension, the proposition containing every world just in case the intension of ϕ includes at least one world, and has the null set as its intension otherwise. More complicated rules involving accessibility relations need to be invoked if more complex notions of necessity and possibility are in play.

Obviously the simple theory presented so far barely scratches the surface of the complexity of natural language: it does not even handle tense or the sorts of sentences we would represent in predicate calculus with multiple variables. The interaction of the syntax and semantics has not been addressed either. But hopefully it gives the flavor of how such a theory might be set up. To see how a possible worlds semantics might earn its keep, we should discuss some more difficult constructions that possible worlds semantics has had some success with.

Some Applications of Possible Worlds Semantics for Natural Language

Possible worlds semantics is useful, as we have seen above, for handling sentential adverbs like "necessarily" and "possibly." It is useful not just for an "unrestricted" sense of necessity and possibility, corresponding to truth in every possible world or some possible world, respectively. It is also useful when dealing with a restricted necessity, such as saying what is necessary given the laws of nature, or what is feasible given some other constraints. (This corresponds to a restriction on the set of worlds relevant for the semantic value.) As was noted in the section above on logic, possible worlds semantics has also been found very useful in dealing with the so-called deontic modalities such as "It is obligatory that . . ." or "It is permitted that . . ." and their more natural-sounding English relatives, and also for dealing with so-called epistemic modalities used for specifying what is known or what is justifiably believed, or what "has to be" or "might be" in the epistemic sense of those expressions. (See Kratzer 1981 for a possible worlds treatment of "ought.")

Possible worlds semantics come into their own for a range of nonextensional uses of language. For example, consider so-called intensional transitive verbs. Examples include "seeks," "owes," "wants," "offers," and a number of others. Whether "John seeks a city of gold" is true or not does not just depend on the extensions of "John" and "a city of gold." Suppose, as is plausible, that there are no cities of gold, and no flying horses. "A city of gold" and "a flying horse" would thus plausibly have the same extension, but "John seeks a city of gold" can be true while "John seeks a flying horse" is false, so "seeks" is intensional. If instead of extensions we use possible world intensions, we can draw the distinction we need. There are possible cities of gold, and possible flying horses, so "a city of gold" and "a flying horse" have different intensions—you could see those intensions as being *the property of being a city of gold* and *the property of being a flying horse*, respectively. Since the intensions are different, if we treat the intension of "seeks" as, in this case, a function from the two intensions that flank it to a proposition, then we can get the right answer, in this example, that the first sentence has the truth-value True at the actual world and the second does not.

One of the applications that caught the attention of the philosophical community following Stalnaker 1968 and Lewis 1973 was the use of possible worlds semantics to give a theory of the truth-conditions for conditional sentences. According to Stalnaker's proposal, for example, "if A then B" is true at a world *w* just in case a "selection function," which is a function from propositions and worlds to worlds, when it takes as arguments the proposition associated with A and the world *w*, yields a world *x* that is a member of the proposition associated with B. The effect of this is that the conditional is true at a world if the "closest" world to *w* where *A* is true is also one where *B* is true. This formal proposal yields a variety of interesting logics of conditionals depending on what constraints are put on the selection function, and deals with a number of problems that faced traditional theories of conditionals. For example, for "if A then B" to be true seems to require more than that either A is false of B is true (the "material conditional" account): I am currently sitting down, but it does not seem true that "if I had stood up, I would have exploded." On the other hand, requiring that for "if A then B" to be true B has to be true in *every* possible world where A is true is too strong a requirement: "If I had stood up, I would *not* have exploded" seems true, though there is nothing absolutely impossible about a situation where I stand up and then explode. A lot of contemporary work in exploring the semantics of conditionals employs some relative of this approach.

A third important application of possible worlds semantics is in the semantics of *propositional attitude* ascriptions. Sentences like "John believes the number of planets is 10" or "Mary desires that everyone love someone" seem to require a relationship between a person and a proposition: The propositions expressed by "the number of planets is 10" and "everyone love someone," respectively. Constructions like "believes that" are not extensional: substituting one true sentence for another in the scope of "believes that" does not, in general, preserve truth-value. Furthermore, the behavior of propositional attitude contexts is too complex to just treat "that ϕ" constructions as, in effect, referring to propositions. "There is someone such that everyone believes that he is the murderer" cannot be treated as referring to a proposition with "that he is the murderer," since "he" is in effect bound by "There is someone." Moving to a possible worlds semantics allows phenomena like this to be dealt with better than existing extensional frameworks.

The above only scratches the surface of work done in the possible worlds tradition. Using possible worlds semantics arguably illuminates topics in semantics as diverse as adverbs, progressives, dispositional expressions, the semantics of embedded questions, and plays a vital role in theorising about effect of context, including in the development of dynamic semantics. For an introductory discussion to more applications than could be discussed here, see Partee (1989), especially section III.

The Problem of Necessary Co-Extension

Some sentences are not true at any possible worlds, and some predicates are true of any possible objects ("is round and not-round" might serve as an example). Furthermore, some pairs of sentences are true in some possible worlds but are true in all the *same* possible worlds, and some pairs of predicates hold of exactly the *same* possible objects. Since orthodox possible worlds semantics takes sentences true at the same worlds to have the same intensions, and predicates true of the same possible objects to share intensions, this leads to some surprising results.

The possible worlds semantics treatment of intensional transitive verbs, for example, misfires when impossibilities get involved, or at least the straightforward version of possible worlds semantics does. Suppose Alan is seeking a proof of the rationality of π. Such a proof is impossible, so the set of objects, across possible worlds, that meet the condition of being a proof of the rationality of π is the null set. Now, suppose Ben is seeking a round square: Again, given the impossibility of round squares, the condition of being a round square corresponds to the null set. "Proof of the rationality of π" and "round square" have the same possible worlds intension. So we should be able to conclude, by substitution of equivalences, that Alan is seeking a round square, and Ben is seeking a proof of the rationality of π. But that substitution is not valid: Alan seems well within his rights to deny that he is seeking a round square, his behavior is not that of someone who is seeking a round square, and our psychological insight has failed if we report that Alan is doing that. Similar problems arise for verb phrases such as "prays to," "try to," "wish for," "offer," and even "owe"—I might owe you a proof of the rationality of π, if I promised to deliver one.

Propositional attitude ascriptions also are dealt with badly by the basic possible worlds semantic treatment: anyone who believes one strictly impossible thing (a thing true in no possible world) believes them all, for example. As well as the problem of impossible contents, propositional attitude ascriptions have problems with necessarily true contents: Anyone who believes one necessarily true thing believes them all, according to this model, but even if mathematical truths are necessary, they are not that easy to discover! There are also problems with beliefs about contingent matters where what seem to be two different belief contents obtain in the same possible worlds. Suppose Robin Hood and Robin of Locksley are the same person, though few suspect this. The Sherriff of Nottingham can believe that Robin Hood robbed the abbot, while not believing that Robin of Locksley robbed the abbot. Or so we would ordinarily think. If "Robin Hood" and "Robin of Locksley" are *rigid designators*, and pick out the same individual across possible worlds, then the two names have the same possible worlds intension. So, according to the straightforward possible worlds semantics for names at least, "The Sherriff believes Robin Hood robbed an abbot" is necessarily equivalent to "The Sherriff believes Robin of Locksley robbed an abbot": the Sherriff cannot believe one without believing the other. A similar problem will arise for natural kind terms, given the usual theory of how they work—anyone who believes there is water in their glass also believes there is H_2O in their glass, according to a straightforward possible worlds semantics. But not everyone believes that water is H_2O.

Finally, the orthodox possible worlds treatment of conditional statements gives results that many find odd. Conditionals with necessarily false antecedents that share a consequent must all be treated the same, since those antecedents all share an intension: the standard line (for example Lewis 1973: 24–26) is to treat all conditionals with necessarily false antecedents as true at all worlds, regardless of the consequent. However, many people give different verdicts on different "counterpossible" conditionals: "If 8 had been divisible without remainder only by 1 and itself, it would have been prime" seems good, but "If 8 had been divisible without remainder only by 1, 2, 5 and itself, it would have been prime" looks less appealing: in the latter case it is tempting to say it would not have been prime, but it would have been composite instead. (See Nolan (1997) for a case for positing counterpossible conditionals with non-trivial truth values.)

There are a number of responses available to advocates of possible worlds semantics here. The most straightforward is to bite the bullet: Alan *is* seeking a round square,

though he might not put it that way; the Sherriff *does* believe Robin of Locksley robbed an abbot; everyone always knew that water was H_2O, and so on. Biting these bullets often goes with a strategy of explaining the *pragmatics* of these utterances so that they are somehow misleading or incorrect to say, even if they are not false (this seems to be the proposal about some cases in Stalnaker 1978, for example).

Another relatively straightforward response is to say that possible worlds semantics is a *model* of the meanings of linguistic expressions that is useful for some purposes but not others: perhaps it is analogous to Newtonian mechanics, which is a good enough approximation when it comes to bridge-building but not the right physics to use when designing particle accelerators. If this approach is taken, it leaves open the question about what the more fully accurate theory might look like—but for some projects, it makes sense to put that question off.

Another response is to say that a semantics of natural language should use constructions from possible worlds as *one* of the components of a "meaning": perhaps semantic values should be pairs of possible worlds constructions and something more linguistic, for example, or that possible worlds intensions should be components in a "structured meaning" that looks rather like a syntactic structure, as in Creswell (1985). There are potentially as many different versions of this strategy as there are components we could add to a semantic value, so it is difficult to generalize about the prospects of this way of developing possible worlds semantics.

A way of complicating the postulated intensions without bringing in things besides constructions from possible worlds is to treat intensions as "two-dimensional intensions." David Kaplan pioneered this move to deal with demonstratives and indexicals: expressions like "that" or "I" or "tomorrow" that make different contributions to truth-conditions on different occasions of use (Kaplan 1989). "It will rain tomorrow" can be true when I say it today but false when I say it tomorrow, for example. Kaplan held that the meaning of an expression like "that" or "tomorrow" had two aspects: a *content*, which was in effect a standard possible worlds intension, and a *character*, which was a function from certain n-tuples (with components like the speaker, and the time of utterance) to contents. Such n-tuples can be generalized to *worlds* of utterance, or worlds of utterance plus a "center" to indicate which speaker, time, audience etc. in the world is being picked out—these are sometimes called the 'context' of the utterance.

While Kaplan intended his framework for a limited range of context-dependent expressions, later "two-dimensional" theorists have offered similar treatments for proper names and natural kind terms. The *contents* of proper names pick out the same object in each world, like a standard possible worlds intention, but the character associated with a name maps the name to different contents depending on features such as which object is at the source of the reference-chain associated with the indicated speaker, or the referential intentions of the designated speaker. In this framework, there are contexts which take "Robin Hood" and "Robin of Locksley" to different contents, and since the overall functions-from-contexts-to-content intensions are thus different for the two names, there is scope for those intensions to interact differently with pieces of language such as "believes that." (See Garcia-Carpintero and Macia (2006) for a collection containing a number of papers for and against such generalized two-dimensional treatments.)

A final kind of modification for orthodox possible worlds semantics to deal with the problem of necessary co-extension is to be more generous with what worlds the theory allows. We got the problems because some expressions had equivalent applications in

every possible world: "a proof of the rationality of π" and "a round square", for example, or "Robin Hood" and "Robin of Locksley". If we allowed worlds to have the extensions of these expressions come apart from each other, these problems would not arise.

One way to do this, which is tempting in the case of propositional attitude ascriptions and intensional transitives involving the activity of agents, is to use *epistemically* possible worlds or *doxastically* possible worlds. These worlds might be more generous than standard possible worlds: For example, if it is possible to believe ϕ without believing φ, there are "worlds" where ϕ and φ come apart, and connectedly that the extensions of two expressions (like "water" and "H_2O" or "proof of the rationality of π" and "round square") come apart at some world if it is possible to believe that an object is in one extension but not in the other. Or one could go even further and drop the restriction to do with our capacities to believe: If a worlds semantics allowed arbitrary impossible worlds as well as possible worlds, the problem of necessary co-extension could be guaranteed to not arise.

Possible worlds semantics has provided a powerful and impressive framework for systematically dealing with the semantics of puzzling constructions in language. Whether the theory of semantics it yields needs to be extended by other devices, and if so what those devices are, will continue to be one of the important debates in the philosophy of language in the coming decades.

Related Topics

1.1 Extensions, Intensions, Character, and Beyond
3.1 Names
5.4 Modal Logic
5.9 Montague Grammar
6.6 Necessity and Meaning
6.7 Propositional Attitude Reports.

References

Copeland, B.J. 2000. "The Genesis of Possible Worlds Semantics," *The Journal of Philosophical Logic* 31: 99–137

Cresswell, M. 1985. *Structured Meanings*. MIT Press, Cambridge MA.

Garcia-Carpintero, M. and Macia, J. (eds.) 2006. *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press, Oxford.

Kaplan, D. 1989. "Demonstratives," in Almog, J., Perry, J., and Wettstein, H. *Themes from Kaplan*. Oxford University Press, Oxford. 481–563.

Kratzer, A. 1981. "The Notional Category of Modality," in Eikmeyer, H.J. and Rieser H. (eds.) *Words, Worlds, and Contexts*, de Gruyter, Berlin: 38–74.

Kripke, S. 1959. "A Completeness Theorem in Modal Logic," *Journal of Symbolic Logic* 24: 1–14.

Kripke, S. 1963. "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* 16: 83–94

Lewis, D. 1973. *Counterfactuals*. Blackwell, Oxford.

Lewis, D. 1986. *On the Plurality of Worlds*. Blackwell, Oxford.

Montague, R. 1973. "The Proper Treatment of Quantification in Ordinary English," reprinted in Montague 1974, 247–270.

Montague, R. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. (Thomason, R. ed.) Yale University Press, New Haven.

Nolan, D. 1997. "Impossible Worlds: A Modest Approach," *Notre Dame Journal for Formal Logic* 38.4: 535–572

Partee, B. 1989. "Possible Worlds in Model-Theoretic Semantics: A Linguistic Perspective," in S. Allen

(ed.) *Possible Worlds in Humanities, Arts and Sciences. Proceedings of Nobel Symposium 65*. Walter de Gruyter, New York, 93–123.
Routley, R., Meyer, R.K., Plumwood, V. and Brady, R. 1983. *Relevant Logics and its Rivals*, Volume I. Ridgeview, Atascardero, CA.
Stalnaker, R. 1968. "A Theory of Conditionals," in Rescher, N. (ed.) *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, no.2. Basil Blackwell, Oxford., 98–112
Stalnaker, R. 1978. "Assertion," in Cole, P. (ed). *Syntax and Semantics: Pragmatics* Vol. 9. Academic Press, New York, 315–332

2.7
DYNAMIC SEMANTICS

Seth Yalcin

1 Introduction

'Dynamic semantics' gets used in various, not always compatible ways. On perhaps its most inclusive use, it applies to a range of semantic systems embracing a thesis we might call:

> DISCOURSE PRIMACY. It is fundamentally entirely discourses that have truth-conditions (or more broadly, informational content). Individual sentences have truth-conditions in at best a derivative sense, insofar as they have some potential to impact the truth-conditions of a discourse.

This thesis breaks with a long semantic tradition—going back at least to Frege, and running through Montague and beyond—which revolves around the individual sentence, articulating the meaning of all linguistic expressions in terms of their contributions to the truth-conditions of a sentence. By shifting the locus of truth-conditions in systematic theorizing to the discourse, DISCOURSE PRIMACY shifts the traditional center of gravity in semantics. It recommends that we articulate the meaning of a sentence in terms of its potential to contribute to the truth-conditions, or informational content, of discourses in which it can occur. This idea was first developed by Kamp [1981] and (independently) by Heim [1982], in works that formed the starting point for subsequent theorizing in the dynamic tradition.

Exactly how to model a discourse is itself a theoretical question on which theorists working in the spirit of DISCOURSE PRIMACY may differ. But a common theme in the dynamic tradition is to model a discourse via aspects of the mental states of the agents in conversation. For instance, Kamp [1981] models a discourse via a certain kind of a structured representation (a *discourse representation structure*, or DRS), which (he postulates) is an abstract model of the kind of mental representation involved in linguistic processing and understanding. On his picture, it is fundamentally these structures that have truth-conditions. And in Heim [1982], the fundamental bearers of truth-conditions are the coordinated states of presupposition of the participants of the conversation. It is the content of these states that the meanings of sentences impact.

Once we take the dynamic turn and accept DISCOURSE PRIMACY, we come to a fork in the road concerning how to articulate the meanings of sentences. One path accepts a thesis we can call DYNAMIC REPRESENTATION; the other, a thesis we can call DYNAMIC INTERPRETATION.

DYNAMIC REPRESENTATION. Sentences encode instructions for updating a certain kind of representation, and it is this representation, not sentences or other linguistic expressions, that is the primary object of compositional semantic interpretation.

DYNAMIC INTERPRETATION. The compositional semantic value of a sentence is a function from a body of information to a body of information.

The first thesis we find in Kamp [1981]; the second in Heim [1982]. The theses are obviously different. As noted, Kamp's account centrally involves a certain intermediate layer of representation, a DRS. He supplies a systematic construction procedure for mapping a sentence or sequence of sentences to a DRS, and a systematic procedure for interpreting DRSs. The resulting system is not locally compositional in the usual sense, for it does not directly associate linguistic expressions with interpretations. (For relevant discussion see Groenendijk and Stokhof [1991], Muskens [1996], Muskens et al. [2011], van Eijck and Kamp [2011].)

We don't directly discuss DYNAMIC REPRESENTATION here. It is better treated against the larger context of Discourse Representation Theory (DRT), the family of approaches to semantic theorizing—some straightforwardly compositional, some not—that followed in the footsteps of Kamp [1981]. For introductions to DRT, see Kamp and Reyle [1993], Geurts and Beaver [2007], van Eijck and Kamp [2011].

On a more restrictive use of 'dynamic semantics', it applies just to those accounts that embrace DYNAMIC INTERPRETATION. This restricted use is the one in play henceforth. This article is about the import and upshot of DYNAMIC INTERPRETATION—for short, DYNAMIC. It is about the idea that the compositional semantic value of a sentence is the sort of thing that updates a body of information—often, though not always, by adding to it more information. In the jargon, it is about the idea is that the semantic value of a sentence is a *context-change potential*, or *update function*, or *program* on a body of information.

To get a handle on DYNAMIC, contrast it with its much more traditional cousin, STATIC:

STATIC. The compositional semantic value of a sentence is a truth-condition—a function from possible worlds to truth-values, or more generally, a function from indices (points of evaluation with a model) to truth-values.

We can think of STATIC and DYNAMIC as competing high-level empirical hypotheses about what kind of thing the semantic value of a sentence is.

We have two aims. First, we should like a basic sense of how, in principle, DYNAMIC could be incorporated into a theory which performs the sort of explanatory work traditionally expected from STATIC.

This we tackle in section 2. Second, we should like a sense of some of the specific linguistic phenomena that have been taken to recommend DYNAMIC over STATIC. This we broach in section 3.

2 Dynamic Semantic Values

How, in principle, might dynamic semantic values be put to work in explaining the kind of phenomena we have always expected our semantic values to cover?

That prompts the prior question: what is it we expect semantic values to explain? Theorists of course differ in exactly what work they expect this notion to perform, but let us focus on a few of the most usual explananda.

First there is the *productivity* of language use: the empirical fact that competent speakers can understand and produce complex expressions that they have never before encountered. This is a typical motivation—perhaps the leading motivation—for embracing compositionality. [See entry on COMPOSITIONALITY.] The assumption of compositionality is a point of common ground between STATIC and DYNAMIC. Typical defenders of each assume compositionality partly with an eye towards explaining the productivity of language use

More illuminating differences between the sort of theory one gets out of DYNAMIC as opposed to STATIC emerge when we look at other explanatory demands we typically place on semantic values. Besides helping to explain productivity by being compositional, the semantic values of sentences are normally expected to do the following explanatory work:

(i) The semantic values of sentences should play a central role in explaining how speakers *communicate*, and in particular, *transfer information*, by using the sentence.

(ii) The semantic values of sentences should play a central role in explaining which sentences follow from which, and which sentences are incompatible with which.

These tasks will be approached in importantly different ways, depending on whether one embraces STATIC or DYNAMIC.

2.1 Communication

Take first (i). How, assuming static, is meaning leveraged to communicate information? A familiar and traditional story is the following. By STATIC, sentences have truth-conditions. We take it these truth-conditions constitute a core aspect if not the core aspect, of the informational content normally communicated in context by an assertion of the sentence. Building on Grice and Stalnaker, we take it communication takes place against a background of mutually shared presuppositions—a *common ground*—and that the characteristic pragmatic impact of an assertion is to add the informational content determined by the sentence uttered to the common ground of the conversation. Thus an assertion normally serves to update the common ground, partly as a function of the truth-conditional content of the sentence asserted in context. It is a feature of the pragmatics of assertion that the common ground is normally to be updated in this way. (See Stalnaker [1975], [1978].)

This story assumes STATIC, but note that on it an asserted sentence is nevertheless associated with a rule for updating the common ground—in effect, with what we have called an update function, or context change potential. How then does DYNAMIC differ from STATIC in respect of (i)? The core difference is that on a dynamic view, the context change potential of a complex sentence is determined compositionally by the semantic values of its parts. Ancillary pragmatic assumptions are not required. And correspondingly, the compositional contribution of an *embedded* sentence—the semantic ingredient the sentence contributes to the larger linguistic expressions in which it occurs—is

itself taken to be a context change potential. Thus the context change potential of a conjunction, for instance, is a literally function of the context change potentials of its parts. This is quite different than the static version of the story, where the context change potential of a sentence is only determined with the help of pragmatics, and applies only in connection with *unembedded*, asserted sentences.

When it comes to conversational update, the dynamic view rolls into the semantics what, on a static view, would be parceled to pragmatics. The compositional contribution of a sentence is just identified with an instruction for updating the informational context, or common ground, of a conversation.

2.2 Consequence

Turn next to (ii), the demand that the semantic values of sentences play a central role in explaining which sentences follow from, or are consequences of, which. On a classic static view, the story is straightforward: sentences have truth-conditions, and consequence is a matter of truth-preservation. An argument is valid, we can say, just in case if the premises are true, the conclusion must also be true.

Since a dynamic semantics does not generally associate sentences with truth-conditions, it cannot tell the same story. (It might try to tell the same story indirectly, by associating all declarative sentences with context-change potentials which determine, in some systematic way, truth-conditions. But this would be a severe constraint, and one precluding many of the motivating applications for dynamic semantics in the literature.) To cover the entailments evidenced to us by the productions of competent speakers, and so perform the explanatory work required by (ii), dynamic semantic systems are typically equipped with alternative formalizations of consequence.

Central to most of these formalizations is the notion of a *fixed point* of a sentence. The fixed points of a sentence will be those bodies of information which, when operated on by the update function semantically expressed by the sentence, return just that body of information back. Intuitively, the fixed points of a sentence will be those bodies of information that already incorporate the informational update associated with the sentence. The fixed points of a sentence ϕ are said to *support* ϕ, or *accept* ϕ.

Generally speaking, the update function of a sentence will map bodies of information to fixed points of that sentence. For example, on the dynamic model, when an assertion gets accepted in conversation, the common ground of the conversation normally becomes a fixed point of the sentence. (That is why we generally don't say things twice: once our conversation is fixed point of a sentence, uttering the sentence will not—at least as a matter of the semantics of the sentence—alter the information that is common ground.)

Once we have the idea of a fixed point of a sentence, we can define consequence in terms of it. Here there are various options, of which we will just mention one (for additional discussion and further options, see for instance van Benthem [1996], Veltman [1996], Beaver [2001], Muskens et al. [2011] and references cited therein). Consequence may be defined as follows: A conclusion ψ follows from some premises $\phi_1, \ldots, \phi_n$ just in case every body of information which is a fixed point of the premises is also a fixed point of the conclusion. On this analysis, consequence is the sort of thing that preserves acceptance. Alternatively put, consequence is an informational update-

preserving relation between sentences. The thought is that once a state of information is updated with the premises of a valid argument, it already incorporates the update associated with the conclusion.

The idea that the data of semantics largely concerns the truth-conditions of sentences is sometimes treated as a platitude. But in a context in which we are debating the virtues of STATIC versus DYNAMIC, we must acknowledge that this is really a theory-laden and question-begging way to describe the matter. The dynamic approach offers a rival characterization of the data of semantics: The data are, at the most general, data about the potential of a sentence to update or change a conversation. It is fundamentally these kinds of facts that judgments about consequence and compatibility track.

Lewis has famously written that "Semantics with no treatment of truth-conditions is not semantics" ([Lewis, 1970, 18]). What he meant to underscore in saying this was that a semantics should associate expressions with *interpretations* as opposed to mere translations. It is worth emphasizing that in this respect, dynamic semantics and truth-conditional semantics are on the same side of the fence: They are each card-carrying members of the interpretive (representational, model-theoretic) tradition in semantics. (But there is the related question of whether the role of an intermediate layer of representation is doing nontrivial explanatory work; we return to this question, in the context of Heim's account, briefly in section 3.3.)

Still, dynamic semantics is disruptive to traditional distinctions in other ways. Notably, talk of sentences expressing propositions, or having propositional content, often lacks clear sense in a dynamic setting, as does the traditional distinction between a sentence's informational content and its force. The reason, again, is that the chief locus of informational content on a dynamic picture is not the sentence but rather the states of presupposition of the participants in the discourse. It is the content of these states that fixes the common ground, and sentence meanings are foremost tools for manipulating this shared body of informational content.

2.3 A Sample Semantics: Negation and Conjunction

The preceding gives an abstract idea of how DYNAMIC might in principle be worked into a theory which explains the sort of things we expect a compositional semantics to explain. But a real sense for the dynamic approach can only come through working with concrete examples. For illustrative purposes, then, let us consider an example of a toy dynamic semantics, one defined in an ordinary artificial propositional language L containing just negation and conjunction. We will soon turn to more complicated systems.

Defs. A **model** M for L is a triple $\langle W, I\rangle$ where W is a set of possible worlds, and I is an **interpretation function** mapping the propositional letters of L to sets of worlds.

Def. A **context set** in M is any subset of W_M.

Def. For any M, a **dynamic valuation** [ ] (for M) is a function from wffs of $L$to functions from context sets (in M) to context sets (in M) recursively defined as follows, where α is any propositional letter, ϕ and ψ are any wffs, and c is any context set in M:

$c[\alpha] = c \cap I(\alpha)$
$c[\neg\phi] = c - c[\phi]$
$c[\phi \wedge \psi] = c[\phi][\psi]$

Def. If $\phi$is a wff, $[\phi]$ is the **context change potential** (or **update function**) of ϕ.

Note we are employing *postfix* notation: the argument of context change potential is written to its left, rather than its right. This allows us to use the left–right direction to reflect the order in which a sequence of sentences updates a body of information.

Central to any dynamic semantics is a stand about how to formally model the informational context of a conversation—the object that dynamic semantic value act upon. The above semantics encodes the idea of representing the informational common ground of a conversation via a *context set*, the set of possible worlds left open by what is mutually presupposed by the interlocutors, an idea first developed in detail by Stalnaker [1975, 1978]. Information growth in conversation is a matter of eliminating possibilities from the context set. Sentences characteristically act to knock worlds out of context sets, thereby growing the amount of information taken for granted in the conversation.

This semantics illustrates the way in which the context change potential of a complex sentence may be defined in terms of the context change potentials of its constituent parts. Now we can formalize the dynamic notions of acceptance and consequence informally described above:

Def. c **accepts** ϕ just in case $c\,[\phi] = c$.

Def. A sentence ϕ is a **dynamic consequence** of a set of sentences Γ, $\Gamma \models_i \phi$, just in case for any c, whenever all of the sentences of Γ are accepted in c, so too is ϕ.

And from here, it is not hard to see that this semantics validates the same patterns of inference as would an ordinary classical semantics defined on the same language. Thus $(\phi \wedge \psi) \models \phi$ is validated, $\neg\neg\phi \models \phi$ is validated, and so on. (See van Benthem [1986, 1996], Groenendijk et al. [1995], von Fintel and Gillies [2007], and Muskens et al. [2011] for more discussion and relevant results.)

This toy dynamic semantics has two notable properties: it is *eliminative* and *distributive*.

Def. A dynamic valuation function $[\cdot]$ is **eliminative** just in case $s[\phi] \subseteq s$, for all s, ϕ.

Def. An dynamic valuation function $[\cdot]$ is **distributive** just in case $s[\phi] = \cup_{i \in s}\{i\}[\phi]$, for all s, ϕ.

Distributivity says that the output context of any update just as well might be arrived at by updating on singletons in the prior context and aggregating the results. Eliminativity entails that this aggregate set will always be some subset of the original context.

These two features entail that our toy dynamic valuation function acts much like a sieve. If we run a bucket of flour through a sieve, the output is a new, smaller quantity of flour from that bucket. Which particles the sieve lets through is entirely a matter of the individual particles and their sizes. If we had sieved half of the bucket first and then the second half, we still would have gotten the same output flour.

If your dynamic semantics is eliminative and distributive, it means that it is boring enough to easily implement statically (van Benthem [1986]). More precisely, it means that each sentence ϕ corresponds to a static condition, P_ϕ, on points, such that:

$c[\phi] = c \cap P_\phi$ (cf. the assertion rule of Stalnaker [1978])

for any context c. It means, in other words, that we might as well have an ordinary static compositional semantics together with the above simple rule for mapping the static condition a sentence expresses into its context change potential.

What this suggests is that if there is a reason to prefer DYNAMIC to STATIC, it is because there is reason to believe in more sophisticated updates on contexts, ones that go beyond the mere sieving we find in eliminative, distributive dynamic systems. And indeed, the dynamic approaches motivated in the literature typically lack eliminativity, distributivity, or both.

Is the failure of eliminativity and distributivity necessary and sufficient to demonstrate the need for dynamicness? We return to that question briefly below (section 6).

* * *

Now that we have a sense of how dynamic semantic explanations are possible, we can ask what might make explanations in this style desirable. As competing hypotheses about a certain class of semantic values, we should favor DYNAMIC over STATIC only if doing so makes for an overall more explanatory theory. A great many motivations have been offered for dynamic analyses. We will review just three—our main objective being only to get acquainted with some early and/or well-known dynamic semantic accounts, and to achieve a rough sense of some of the terrain. The motivations we review stem from: (i) intersentential and donkey anaphora, (ii) the problem of presupposition projection, and (iii) the semantics of epistemic modals. We proceed in that order.

3 Intersentential and Donkey Anaphora

3.1 Indefinites and their Discontents

We are familiar with the idea of using the apparatus of quantification to model certain kinds of anaphoric reference in natural language. In particular we are familiar with the idea of giving semantics for

(1) A car drove by and it honked.

in a manner which treats the pronoun *it* as a variable bound by an existential quantifier introduced by the indefinite *a car*, so that we may associate the sentence with truth-conditions along the lines of:

(2) $\exists x(x \text{ is a car} \wedge x \text{ drove by} \wedge x \text{ honked})$

What to say, however, when we break (1) into two sentences?

(3) A car drove by. It honked.

Naively, it would seem desirable to semantically associate this brief discourse with the same truth-conditions as (1) (cf. Geach [1962]). Yet this would apparently require the scope of the indefinite in the first sentence to reach across the sentence boundary so as to bind *it* in the second sentence—not an option, given ordinary variable binding is a strictly intrasentential affair.

Should we instead understand the semantic relation between *a car* and *it* in (3) on the model of coreference? That would call for the availability of a nonquantificational, referential interpretation of indefinite noun phrases. But this idea quickly runs into difficulties. For instance, it raises the question why discourse:

(4) The man didn't buy a car. # It was black.

is defective (cf. Karttunen [1976]). If the indefinite and the pronoun were understood on the model of coreference, we would be able to hear (4) as saying that a specific car, which the man didn't buy, was black. But it is not easy to hear it that way.

So two obvious ways to construe the semantic relation between the indefinite and the pronoun in (3) face clear obstacles. This puzzle has an intrasentential analogue in the phenomenon of donkey anaphora (Geach [1962]). Consider:

(5) Every farmer who owns a donkey beats it.

It is obvious that the indefinite in this sentence is not a referring expression. But nor can we straightforwardly understand the anaphoric pronoun *it* as bound by the indefinite, for the pronoun is not within the indefinite's scope.

Moreover, if we wished to render the truth-conditions of the natural reading of (5) in the language of predicate logic, we would reach for the following:

(6) $\forall x \forall y\, [(\text{farmer}(x) \wedge \text{donkey}(y) \wedge \text{owns}(x, y)) \rightarrow \text{beats}(x, y)]$

Somehow, the deeply embedded indefinite in (5) appears to really take wide scope over the whole sentence—and moreover express universal, rather than existential, quantification. It is hard to understand how these truth-conditions could be achieved compositionally.

It was in large part to address these and related puzzles about intersentential and donkey anaphora that Kamp [1981] and Heim [1982] each proposed their dynamic semantic systems. Indeed, the semantics of anaphora is arguably the leading historical motivation for going dynamic (with presupposition projection coming in at a close second). We will focus on the *file change semantics* developed in Heim [1982], but we emphasize that the overlap between Heim's and Kamp's accounts is quite substantial.

3.2 File Change Semantics

In its treatment of anaphora, file change semantics weaves together four distinct innovations. First, it employs a novel representation of context, one which can be seen as enriching the Stalnakerian context sets used in our toy semantics above. Second, it treats indefinite noun phrases (*a cat*) as semantically akin to open sentences (*x is a cat*),

and not as introducing any quantificational force of their own. Third, building on Lewis [1975], it incorporates the idea of unselective quantification, allowing quantifiers to bind (not just one but) all variables free in their scope. Fourth and finally, file change semantics embraces DYNAMIC.

Beginning with the first innovation, Heim recommends that we take the information presupposed in a conversation to have a structure analogous, in important ways, to that of a filing cabinet. The metaphor goes back at least to Karttunen [1976: 364]:

> Consider a device designed to read a text in some natural language, interpret it, and store the content in some manner, say, for the purpose of being able to answer questions about it. To accomplish this task, the machine will have to fulfill at least the following basic requirement. It has to be able to build a file that consists of records of all the individuals, that is, events, objects, etc., mentioned in the text and, for each individual, record whatever is said about it.

Inspired by this metaphor, Heim models the common ground of a discourse as what we can call a *Heim context*. Heim contexts are defined in terms of several other notions. I will now barrage you with definitions, and then explain the intuitive idea.

> Def. An **assignment world** is a pair of a possible world and a variable assignment.
>
> Def. A **satisfaction set** is a set of assignment worlds.
>
> Def. A **domain** is a set of numerals, each corresponding to possible variable index.
>
> Def. A **Heim context** c is any pair of a satisfaction set s and a domain d which is such that if $n \notin d$, then the variable x_n is free in c.
>
> Def. A variable x is **free in** c if for any two assignment functions f and f' that differ only in their assignment to x, and any world w if $(f,w) \in c$ so is (f',w).

How is a Heim context anything like a filing cabinet? The elements of a Heim context's domain are in one–one correspondence with the files of the cabinet. The cardinality of this domain gives the number of files, telling us how many distinct individuals are assumed to be tracked in the conversation at a given point. On Heim's intended application, the domain will tell us which individuals are available for anaphoric reference in the conversation. In the terminology of Karttunen [1976], the domain tracks the *discourse referents* of the conversation at a given point. (It will become clearer how the domain does this below.) The satisfaction set corresponds to the informational content of the files taken as a whole, and considered independently of their particular file structure. It imposes properties and relations on the individuals assumed to be tracked in conversation, as well as conditions on the world at large.

Think of Heim as beginning with the Stalnakerian context sets of our toy dynamic semantics, and enriching these context sets in two ways. First, each world in the context set is paired with a variable assignment, yielding an assignment world. Assignment worlds are the sorts of things we can evaluate *open* sentences with respect to. Second, we pair this satisfaction set with a domain. As more individuals are made available for

anaphoric reference in conversation, the domain grows. (New "files" are opened.) As more information become presupposed in conversation, the satisfaction set shrinks.

To get a sense of Heim's system, we can begin with its treatment of donkey sentences such as (5). Ignoring various subtleties, take it the underlying syntax for our donkey sentence (5) is as follows (see Muskens et al. [2011] for a brief summary of Heim's syntactic assumptions):

(5) Every farmer who owns a donkey beats it.
every[x_1 is a farmer $\wedge$ x_1 owns x_2 $\wedge$ x_2 is a donkey][x_1 beats x_2]

Here we observe the second innovation mentioned above: the indefinite noun phrase *a donkey* introduces no quantification on its own. It merely contributes what is in effect an open sentence. (Indeed, Heim takes *all* noun phrases to express open sentences. We return to that idea in connection with definiteness below.)

Now to the semantics. Heim's dynamic semantic values for sentences are functions on Heim contexts. These semantic values must update both the domain of the context (which they will do either by leaving it alone, or expanding it) and the satisfaction set of the context (which they will do either by leaving it alone, or reducing it). It is helpful to start out by ignoring domain dynamics entirely. This will help us see where exactly domains are needed.

Focus first, then, on the way in which the meaning of a sentence changes the satisfaction set. Our donkey sentence is built from atomic (open) sentences, conjunction, and universal quantification, so our objective is to give compositional dynamic semantic values for these linguistic expressions. We take it that, as in an ordinary static intensional semantics, to each n-place predicate F^n of the language there corresponds an intension F, taking n individuals and a possible world into a truth value. Then, if s_c is the satisfaction set of c, we can recursively state the dynamical impact of atomic sentences and conjunctions on satisfaction sets as follows:

$s_c[F(x_i, \ldots, x_n)] = \{(f, w) \in c : \mathsf{F}(f(x_i) \ldots f(x_n), w) = 1\}$
$s_c[\alpha \wedge \beta] = s_c[\alpha][\beta]$

The update rule for atomic open sentences is simple: It takes a satisfaction set and eliminates from it all those assignment worlds not satisfying the open formula. The update rule for conjunction is also straightforward, and indeed just the same as in our toy semantics above: It amounts to consecutive update on the input satisfaction set.

Turning to universally quantified sentences, a dynamic rule sufficient to handle (5) is the following:

$s_c[\textbf{every}(\alpha, \beta)] = \{(f, w) \in c$: for every $f' \sim f$:
if $(f, w) \in s_c[\alpha]$, then $(f, w) \in s_c[\alpha][\beta]$

In effect, a quantified statement takes a satisfaction set and eliminates those assignment worlds which (i) would survive update with the restrictor alone but (ii) would not survive update with the restrictor clause followed by the matrix clause. Observe that this semantics incorporates the idea of unselective quantification: the object language quantifier *every* doesn't bind just one distinguished variable, but rather binds all variables free in its scope. This indiscriminate binding enables the single quantifier *every* to

mimic the truth-conditions achieved with the two selective universal quantifiers in (6), yielding the intuitively correct interpretation for (5). Thus if our initial Heim context is *c*, (5) asks us to eliminate those assignment worlds from c which satisfy

x_1 is a farmer $\wedge$ x_1 owns x_2 $\wedge$ x_2 is a donkey

but which fail to satisfy

x_1 beats x_2

Contexts which accept (5), then, are ones that rule out the possibility of ever including a file on a farmer who is a donkey-owner but not a donkey-beater. Or again, the worlds left open by the satisfaction sets of such contexts are worlds whose donkey-owning farmers are all donkey-beaters.

This buys us an account of an important class of donkey sentences. But to pay for it, we had to give up the idea that indefinites introduce their own quantification. So what do we say of the undeniable existential readings of indefinites in rather more ordinary constructions—for instance in our (1) above (repeated below)?

(1) A car drove by and it honked.

Under Heim's syntactic assumptions, (1) is, at the relevant level of abstraction, equivalent to:

x_1 is a car $\wedge$ x_1 drove by $\wedge$ x_1 honked

We naturally read (1) as an existential claim, however. Why, if it is really an open sentence?

Heim accounts for the existentially closed reading of indefinites in the context of defining truth. In characteristically dynamic fashion, she defines truth in the first instance for contexts, and then derivatively for sentences. We only need her definition of truth for contexts to explain the intuition that sentences such as (1) are existentially closed.

> Def. A Heim context *c* is **true** iff there exists, with respect to the actual world @, some assignment *f* of values to variables—equivalently, some sequence of individuals *f* in @—such that $(f, @) \in s_c$.

This definition tells us how to recover the truth-conditional content of a Heim context from its satisfaction set. In effect, the satisfaction set represents a complex intensional relation, and the definition of truth says that a satisfaction set is true just in case this relation is actually instantiated by some objects. The existential reading of sentences such as (1) can then be explained by the fact that if this sentence is accepted in a context *c*, *c* can be true only if there exists a car that drove by and honked. This definition of truth in effect enables a form of wide scope existential closure at the end of interpretation for the variables left free in sentences.

From here, it is easy to handle our case of intersentential anaphora:

(3) A car drove by. It honked.

On Heim's analysis, this discourse has exactly the same context change potential as (1). For recall that on the dynamic account, conjunction merely amounts to consecutive update. When a conjunction is unembedded, that is equivalent to consecutive assertion. Thus (3) is explained on the same model as (1), with the existential reading again accounted for by the definition of truth for contexts.

We have, then, a single semantics predicting ordinary existential readings of indefinites, donkey anaphoric readings, and intersentential anaphoric readings—a highly nontrivial achievement.

But we are not yet out of the woods. Although Heim's definition of truth enables us to supply the existential readings in cases such as (1) and (3), it cannot supply existential readings where the indefinite is itself *scoped under* other operators. Negation provides a simple example of such a case:

(7) There isn't a cloud in the sky.

Clearly, we do not want (7) to be equivalent to something like:

$\neg$ (x_1 is a cloud $\wedge$ $x_1$is in the sky)

For when we apply the definition of truth to get the existential reading of the indefinite the sentence, the resulting interpretation would require that there exists something which is not a cloud in the sky. But that is wrong. On the natural reading, (7) does not say, of some thing, that it isn't a cloud in the sky; rather it simply says that the sky is cloud-free. What we should like to happen with (7) is that the sentence be existentially closed *before* negation is applied:

$\neg\exists x_1$ (x_1 is a cloud $\wedge$ x_1 is in the sky)

Heim manages this by allowing negation itself to existentially close free variables not in already in the domain. This is where domains enter into the story. The central semantic role of domains is to enable the existential closure of variables in embedded contexts.

To say how this works, we now must explain how sentences dynamically change the domain of a Heim context. If c is a Heim context and d_c its domain, we can state the dynamical impact of sentences on domains recursively via the following schemas:

$d_c[F(x_1, \ldots, x_n)] = \{d_c \cup i, \ldots, n\}$
$d[(\alpha \wedge \beta] = d[\alpha] \cup d[\beta]$
$d_c[\neg\alpha] = d_c$
$d_c[\textbf{every}\ (\alpha, \beta)] = d_c$

Open sentences will expand the domain whenever they include variables which are free in c; otherwise, they will leave the domain alone. A conjunction expands the domain whenever one of its conjuncts does. Negated and universally quantified statements such as (5) will never expand the domain.

Observe these rules track intuition about when anaphora is licensed for subsequent discourse. Atomic open sentences and conjunctions containing indefinites both generally license subsequent anaphora:

(8) A cloud is in the sky. It is fluffy.
(9) It is breezy and a cloud is in the sky. It is fluffy.

These kinds of sentences have the potential to "open new files" or "make new discourse referents available." By contrast, indefinites under negation and within quantifier restrictors generally do not:

(4) The man didn't buy a car. # It was black.
(10) Every farmer who owns a donkey beats it. # It is not happy.

Although negated sentences never expand the domain of context, their semantics—in particular, their impact on satisfaction sets—makes essential reference to the context's domain. On Heim's account, the dynamical impact of negation on satisfaction sets is the following:

$s_c[\neg\alpha] =$
$\{(f, w) \in c$: there is no $(f', w) : f' \sim_{d_c} f$ such that $(f', w) \in s_c[\alpha]$

where

Def. $f' \sim_{d_c} f$ iff for all $n \in d$, $f(x_n) = f'(x_n)$.

A negated sentence changes the satisfaction set by eliminating all those assignment worlds which could verify the open sentence α. The procedure for doing this effectively existentially closes the free variables in α which are not in the domain before the negation. This gets the correct result for (7) above.

Once we add domains, we needn't change the update rules we have defined on satisfaction sets for atomic open sentences or for conjunction. But we do have to change our semantics for universal quantification. As the semantics is now stated, it will get incorrect results for sentences such as:

(11) Every farmer who own a donkey doesn't beat it.
every [x_1 is a farmer $\wedge$ x_1 owns $x_2 \wedge x_2$ is a donkey] [$\neg$ (x_1 beats x_2)]
(12) Every farmer who owns a donkey sells it to a merchant. (Kamp [1981])
every [x_1 is a farmer $\wedge$ x_1 owns $x_2 \wedge x_2$ is a donkey] [x_1 sells x_2 to $x_3 \wedge x_3$ is a merchant]

The sentence (11) illustrates a case where negation should *not* existentially close the variables below it, while (12) illustrates the fact that the universal quantifier should not bind variables introduced by indefinites which appear in their nuclear scope. The adjustment to our semantics required to accommodate such examples involves a considerable complication of our update rule for universal quantification. What we need is for the domain to be locally and "temporarily" updated as interpretation proceeds, so that the right variables are bound in the right way. We leave it to the reader to verify whether the following entry will do the job (Heim [1982, 1983a]):

s_c[every(α, β)α)] = {(f,w) $\in$ c : for every (f',w) such that $f \sim_{d_c} f'$ and (f',w) $\in$ $s_c[\alpha]$, there is some (f'',w) such that $f' \sim_{d_{c[\alpha]}} f''$ and (f'',w) $\in$ $s_c[\alpha][\beta]$}

We will get a bit further into Heim's account when we come to the issue of presupposition projection in the next section. Let me close this section with two questions that Heim's dynamic semantics, as so far presented, raises.

3.3 On the File Metaphor

First, there is the question of what exactly to make of Heim's model of the informational context. In Heim [1982], she suggests that her representation of context can been seen as a linguistically motivated extension of Stalnaker's model of the common ground of a conversation in terms of a set of possible worlds—his context sets. But on Stalnaker's picture, a context set models the *content* presupposed by a given set of interlocutors. In what sense can a Heimian context—a pair of a satisfaction set and a domain—be the content of an attitude like presupposition or belief? Insofar as possible, it would be desirable to independently moor this conception of informational content. (Stalnaker's particular model of content is motivated by a broadly decision-theoretic style of functionalism together with a certain externalist treatment of intentionality. (see e.g. Stalnaker [1984]). Of course, one may not accept Stalnaker's particular model of informational content, but the point is that one has a grip on Stalnaker's conception of information quite independently of its application to formal semantics or to the representation of communication. Not so for Heim's model, unless more is said.)

Reflection on this issue may push us closer to an interpretation of Heim's model which brings it closer to Kamp's—one on which the role of an intermediate representation really is, after all, essential. On this reading, Heim contexts would not be contents (as context sets are for Stalnaker) but rather representations having content. Heim seems to resist this interpretation (see, e.g., Heim [1990b]), but it suits her preferred metaphor: she calls her contexts *files*, and ordinary files are, after all, not contents but bearers of content.

On the other hand, there have been efforts to construe at least the satisfaction set component of Heim contexts as contents. For work construable in this direction, see Ninan [2008, 2010], Cumming [2008].

3.4 On the Need for Dynamics

Second, we noted that Heim's semantics weaves together four distinct ideas. It is clear enough how the first three ideas cooperate: The novel representation of context (first idea) allows for the context to be updated by open sentences, and hence by indefinite noun phrases (second idea). These assumptions in turn facilitate the treatment of donkey anaphora via unselective quantification (third idea). But we can ask: in what way is DYNAMIC—the fourth idea—a key ingredient in the story?

Here it is helpful to separate two questions: one about the conceptual necessity of the dynamic formulation, and another about its comparative elegance and explanatory power.

I. Can the work done by file change semantics—specifically, its treatment of donkey and intersentential anaphora—in principle be done within a static semantics, one defined on exactly the same syntactic structures, and without ad hoc stipulations?

II. Does the balance of evidence support the view that a dynamic semantics for donkey and intersentential anaphora is comparatively more elegant and explanatory than any static alternative?

The first question is not directly addressed by Heim, and to our knowledge has not been definitively settled. (Although Heim [1982] itself contains a static semantics for anaphora covering basically the same data as its file change semantics, the static account there requires a more complex syntax, and a number of additional stipulations.) However, it is clear that which assignment worlds are eliminated by any given sentence will depend crucially on whether the variables in that sentence are tracked by the domain of the relevant file. As a result, no simple static association between sentences and conditions on assignment worlds will mimic the updates supplied by file change semantics.

(Might one sidestep this issue by dispensing with domains? Could we just leave out the complexity introduced by domains, and refer instead everywhere to the set of variables not free in the relevant satisfaction set? Here the question is whether "free" should be taken to entail "not-in-the-domain". Heim thinks not: see [Heim, 1982, 303]. Whether or not Heim is right, it is clear that such an adjustment would not make it any easier to associate sentences with fixed conditions on assignment worlds.)

The second question has been, and continues to be, a matter of lively debate within natural language semantics, and remains open. Perhaps the leading static alternative to dynamic analyses of donkey and intersentential anaphora is the definite description ('d-type') approach to pronouns, which takes anaphoric pronouns to be semantically equivalent to certain definite descriptions (see Parsons [1978], Cooper [1979], Heim [1990b], Elbourne [2005], and references cited therein; cf. Evans [1980], Stalnaker [1998]). Another important static semantics is the variable-free approach developed by Jacobson [1999]. For overviews of recent work on pronominal anaphora together with some cost-benefit analysis, see Elbourne [2008].

Heim herself, in Heim [1982] and Heim [1983b], famously suggested that one reason for favoring the dynamic character of her file change semantics was that it allowed for a compelling approach to the projection problem for presuppositions.

4 The Problem of Presupposition Projection

4.1 From Anaphora to Presupposition

We said it would be nice to be able to read (3) along the lines of (2):

(3) A car drove by. It honked.
(2) $\exists x(x \text{ is a car} \wedge x \text{ drove by} \wedge x \text{ honked})$

Heim's semantics enables us to do that. Now suppose we extended this discourse as follows:

(13) A car drove by. It honked. The car was black.

Intuitively it would seem desirable to analyze (13) along the lines of (14):

(14) $\exists x$(x is a car $\wedge$ x drove by $\wedge$ x honked $\wedge$ x was black)

That would require treating definite noun phrases as, like indefinite NPs, open sentences.

Suppose we did treat them that way. Very well; but what then is the semantic difference between definites and indefinites? The obviously do differ, notably in respect of the conditions under which they license anaphora. For instance, *the car*, unlike *a car*, can be read as anaphoric with a previous indefinite ((15)–(16)), and can license subsequent pronominal anaphora even when under negation ((17)–(18)):

(15) [A car]$_i$ drove by. [The car]$_i$ was moving very fast.
(16) [A car]$_i$ drove by. # [A car]$_i$ was moving very fast.
(17) Bill doesn't have [the car]$_i$. [It]$_i$ is in my possession.
(18) Billy doesn't have [a car]$_i$. # [It]$_i$ is in my possession.

If, following the lead of many generative grammarians, we take it DPs/NPs in logical form are generally marked ± with a definiteness feature, the question at issue becomes: What, semantically, does the definiteness feature contribute?

One way to take this question is: what item of truth-conditional content does the definiteness feature contribute? But examples like (15)–(18) suggest that definiteness features are less content-bearing items than discourse management tools—specifically, devices for coordinating speaker on referents. Taking this attitude toward the data and building on earlier work (notably Karttunen [1968]), Heim enriches her system so as to capture the idea that definites, unlike indefinites, are used to refer to something that is *already familiar* at the current state of the conversation. In contrast to indefinites, which generally serve to introduce *new* discourse referents, definites pick up on discourse referents *already in play*. She does this by defining the notion of *appropriateness* for atomic sentences:

> Def. If c is a Heim context and ϕ an atomic sentence, then ϕ is **appropriate** with respect to c iff for every noun phrase NP$_i$ contained in ϕ:
>
> (i) if NP$_i$ is definite, then NP$_i \in d_c$.
> (ii) if NP$_i$ is indefinite, then NP$_i \notin d_c$.

and by declaring that the context change potential of a sentence is *undefined* on contexts with respect to which it is inappropriate:

> NOVELTY-FAMILIARITY CONDITION. If ϕ is not appropriate with respect to c, the context change potential of ϕ is undefined on c.

Let us walk through how this helps with, e.g., (16). The second sentence of discourse (16) contains an indefinite coindexed with an earlier indefinite. By the semantics above, we know that its index must already be in the domain by the time we attempt to update the context with the sentence. So it is inappropriate. By the novelty-familiarity condition, its update is undefined on contexts which accept the first sentence. That predicts infelicity for the discourse, the correct result. By analogous means we can explain the felicity of (15) and (17), and the infelicity of (18). Observe the novelty-familiarity condition charges the domain feature of context with additional semantic work.

(Of course, we needn't view the novelty-familiarity condition as a special semantic stipulation outside of the compositional semantics proper. We may incorporate the appropriateness requirement directly into the dynamic semantics given above.)

One gloss on the preceding is this: by being defined on some contexts but not others, (in)definite noun phrases *take something for granted* about the preceding discourse–specifically, something about the domain of any context to which they are applied. When what they take for granted is wrong, they are inappropriate and (hence) fail to update the context. This glosses raises the question: if sentences or NPs can take something for granted about the *domain* of the contexts they are applied to, can they not also take something for granted about the *satisfaction sets* of the context they are applied to?

Yes, says Heim [1983b]: in fact, that is just the phenomenon usually called *presupposition*. First, define the notion of *admittance* in terms of the definedness of context change potential, as follows:

Def. c **admits** ϕ just in case the context change potential of ϕ is defined on c.

Then define *presupposition* as a relation between sentences in terms of admittance and acceptance, as follows:

Def. ϕ **presupposes** ψ just in case every context which admits ϕ accepts ψ.

Heim conjectured that this definition of presupposition could enable an advance on the *projection problem* for presuppositions: roughly, the problem of predicting the presuppositions of complex sentences in systematic fashion from the presuppositions of their parts. Presuppositions would be generated wherever we have the relevant undefinedness of context change potential; they would *project* whenever the undefinedness of a part was inherited by the whole.

We can illustrate this approach to presupposition projection with three examples, the latter two from Heim [1983b]: (i) conjunction, (ii) conditionals, and (iii) universally quantified sentences.

Conjunctions. The general contours of this account go back to Stalnaker [1974]. Consider:

(19) (a) The king has a son.
(b) The king's son is bald.

Both sentences presuppose that there is a king, and (19b) moreover presupposes (19a). Now the following conjunction embeds both sentences:

(20) The king has a son and the king's son is bald.

but it presupposes only that there is a king. We should like to predict this fact, given only the meaning of *and* and the presuppositions of the primitive parts of the sentence. And we can, with the dynamic semantics for conjunction sketched already in our toy semantics above. On this view, conjunction is merely consecutive update. Although the context change potential of the second conjunct is undefined on context sets which

don't include the information that the king has a son, the same will not be true for the context change potential of the whole conjunction. The reason is that when the first conjunct updates the input context, it yields an intermediate ('local') context which always admits the second conjunct, and which serves as the input to the context change potential of the second conjunct. In the usual terminology, the second conjunct's presupposition is *locally satisfied* by the first conjunct, preventing the undefinedness from percolating up through the composition, and hence blocking the presupposition from projecting. By contrast, the presupposition of the first sentence (that there is a king) does project, for it is never locally satisfied.

Conditionals. The story is similar with conditionals. Consider:

(21) If the king has a son, the king's son is bald.

This sentence too only presupposes that there is a king, despite the fact that its consequent presupposes additionally that the king has a son. We should like to predict this fact, given only the meaning of *if* and the presuppositions of the primitive parts of the sentence.

We can, with with the dynamic semantics for the indicative conditional offered by Heim:

$$s_c[\phi \rightarrow \psi] = s_c - (s_c[\phi] - s_c[\phi][\psi]$$

To update the context with a conditional $\phi \rightarrow \psi$, we eliminate those possibilities (assignment worlds) which would survive update with ϕ but not with ϕ followed by ψ. According to this rule, a conditional is well-defined on s_c only when both $s_c[\phi]$ and $s_c[\phi][\psi]$ are. Thus since (19), the antecedent of (21), is only admissible on contexts that include the information that there is a king, the whole conditional (21) presupposes that there is a king. By contrast, nothing prevents the conditional from being admissible on contexts that fail to accept that the king has a son. Although the consequent (20) does presuppose this, and hence is only welldefined on contexts accepting it, the antecedent serves to locally satisfy this requirement, and thereby prevents the whole conditional from inheriting this presupposition.

Thus the presuppositions of (21) fall out from the context change potential of the conditional together with the presuppositions of its parts. Generally speaking, whether a presupposition projects in a particular setting is a matter of whether it is locally satisfied. If the presupposition of an expression is not guaranteed to be satisfied by its local context—by the context it is semantically operating on—the presupposition will project.

Universally quantified sentences. Consider the following sentence, together with its Heimian syntactic parse:

(22) Every nation cherishes its king.
every [x_1 is a nation] [x_1 cherishes x_1's king]

This sentence presupposes

(23) Every nation has a king.

Considerations of generality make it very natural to want to trace this presupposition to 'x_1's king'—specifically, to its presupposing:

(24) x_1 has a king.

Of course, that would require making sense of the idea of an open sentence being a presupposition. This is a nontrivial challenge for accounts associating presuppositions only with proposition-expressing elements (e.g., Stalnaker [1974], Gazdar [1979]), but as should be obvious, it presents no difficulty for Heim: Her account allows open sentences to update contexts, and thus enables the definition of presupposition given above to apply generally. Applied together with her semantics, it yields the prediction that (22) does indeed presuppose (23), as the reader may confirm. The reader may also wish to confirm that her semantics correctly predicts that this presupposition of (22) can be cancelled with an appropriate addition to the restrictor clause:

(25) Every nation with a king cherishes its king.
every [x_1 is a nation $\wedge$ x_1 has a king] [x_1 cherishes x_1 's king]

See Heim [1983a], Beaver [2001] for further discussion and additional difficulties.

The preceding is intended to give just a flavor of the kind treatment of presupposition projection possible within a dynamic setting. For a much fuller picture, both of the historical background and of the complex array data modern theories of presupposition projection must be tested on, Beaver [2001] is an especially useful source.

4.2 The Question of Explanatory Power

It might be thought that any linguistic connective can be associated with three items: (1) a truth-conditional contribution (a *content* property); (2) a specification the presuppositions it generates, if such there be (a *presupposition* property); and (3) a specification of its permeability for the presuppositions of its arguments (a *heritage* property). The heritage property of the indicative conditional, for instance, might specify that an indicative conditional $\phi \rightarrow \psi$ (a) presupposes whatever ϕ does, and (b) presupposes $\phi \rightarrow \psi'$, where ψ' is what ψ' presupposes.

Karttunen and Peters [1979] presented a framework in which these three properties were treated as mutually independent and specified separately for each connective. But it is natural to conjecture that the heritage property of an expression can somehow be derived from a specification of its content and presupposition properties. Else we have a theory which

> . . . implies—implausibly—that someone who learns the word "if" has to learn not only which truth function it denotes and that it contributes no presupposition, but moreover that it has the heritage property specified . . . It also implies that there could well be a lexical item—presumably not attested yet—whose content and presupposition properties are identical to those of "if", while its heritage property is different. Heim [1983b]

The worry isn't that such an expression is logically impossible. There is no contradiction in the idea of such a lexical item. Rather, the worry is that an apparently robust

empirical generalization about natural languages—roughly, that heritage properties supervene on content-and-primitive-presupposition properties—is being stipulated rather than predicted.

Gazdar [1979] argued that Karttunen and Peters's theory was subject to this kind of objection, and Heim [1983b] agreed. Her account was intended to satisfy this demand for explanatory adequacy while being superior to Gazdar's account in its empirical coverage. Heim thought she had an explanatory advantage over Karttunen and Peters partly because, given only the context change potential for a connective, one could read off both the content and heritage properties.

But as subsequently noted by Rooth (p.c. in 1987 to Heim, quoted in Heim [1990a]) and Soames [1989], differing specifications of the context change potentials of the connectives Heim discusses—differing, in particular, in the heritage conditions they would impose—could in principle generate the same predictions vis-à-vis the informational contribution of the expression to context. To illustrate, contrast the standard dynamic semantics for conjunction (C1) (repeated from our toy semantics) with the deviant entry (C2), where the conjuncts update the context in reverse order:

(C1) $c[\phi \wedge \psi = c[\phi][\psi]$
(C2) $c[\phi \wedge^* \psi = c[\psi][\phi]$

In the context of Heim's semantics, these will have the same update impact were both defined; they differ only in their heritage conditions. So just as Karttunen and Peters need to stipulate that one among a range of possible heritage properties is the correct one for (say) the conditional, Heim is compelled to stipulate that one among a range of possible context change potentials, all equivalent with respect to their informational impact on context, is the one with the correct heritage condition. Heim conceded this worry, and agreed that the explanatory power of her account had been overstated (Heim [1990a]).

To be clear: it would be incorrect to say, concerning the limited array of projection facts Heim's account was directed at, that the account is completely stipulative, hence not explanatory or predictive in character. On the contrary: on the basis of a finite set of stipulations, Heim's account generates predictions about the presuppositions of an infinite class of sentences. Rather, the point is simply that (i) a theory which got by with less—specifically, which built-in fewer assumptions about heritage conditions—would be *more* explanatory; and (ii) there is good reason to think that an alternative theory, better in specifically this respect, can be had.

This is a worry for an account along Heim's lines, insofar as it shows the account does less than we might have wanted. Is it an *objection* to the account? Not really—not, anyway, until some relevantly more explanatory alternative theory is actually put on the table.

Recent work by Philippe Schlenker has brought renewed attention to this worry about dynamic accounts of presupposition projection, and constitutes an important attempt to deliver the explanatory goods lacking in Heim's account. See Schlenker [2008, 2009], and the critical replies to Schlenker by Beaver [2008], Fox [2008], and Rothschild [2008b].

Another important attempt to fill the explanatory gap in dynamic accounts of presupposition projection, one not discussed by Schlenker, occurs in Beaver [2001]. He writes:

> I will borrow from Veltman's work to show how the context sensitivity of [epistemic modal] words like 'might' and 'must' motivates a dynamic semantics. None of the alternative CCPs for connectives that have been suggested by Rooth and Soames would be compatible with this semantics, and it is hard to imagine how a relevantly different dynamic semantics could still get the facts right about the meanings of the epistemic modalities.
>
> (146)

Beaver is correct: the dynamic semantics for epistemic modals he considers is not compatible with the deviant CCPs offered by Rooth and Soames. Its addition to Heim's system would thus appear to increase the system's overall explanatory power *vis-à-vis* the problem of presupposition projection. So let us have a look at this semantics and its motivation.

5 Epistemic Modals

The dynamic semantics Beaver has in mind is a refinement of the update semantics for epistemic modals developed by Veltman [1996] (a predecessor of which occurs in Veltman [1985]; cf. also Stalnaker [1970], Groenendijk and Stokhof [1975], Groenendijk et al. [1995]). To illustrate the basic idea of the semantics, we can extend our toy dynamic semantics from section 2.3 with an epistemic possibility operator $\Diamond$ (for epistemic *might*, *may*, *possibly*, etc.), adding the following clause:

$$c[\Diamond\phi] = \{w \in c : c[\phi] \neq \emptyset\}$$

(Epistemic *must* can then be introduced as the dual of $\Diamond$.) Given ϕ is not itself modalized, this captures the thought that $\Diamond\phi$ is accepted with respect to a context just in case, roughly, there is a ϕ-world left open by (in) the context. Applied to c, the function $[\Diamond\phi]$ returns either c or the empty set. Veltman uses the metaphor of a test: "Sentences of the form *might* ϕ provide an invitation to perform a test on c rather than to incorporate some new information in it" ([Veltman, 1996, 10]. Sketching roughly the same idea in earlier work, Stalnaker writes: "A sentence of the form 'It may be that P' . . . may be interpreted as making explicit that the negation of P is not presupposed in the context" ([Stalnaker, 1970, 286–7]).

On this account, $\Diamond\phi$-sentences never, as a matter of their semantics, add information to the context when tokened unembedded (where "adding information" corresponds to winnowing down the possibilities, without excluding all possibilities). As a result, even as one strictly increases the information taken for granted by the context, the context can go from accepting $\Diamond\phi$ to failing to accept it. In the jargon, $\Diamond\phi$ is not *persistent*:

> Def. ϕ is **persistent** just in case if c accepts ϕ, then ϕ is accepted in all the subsets of *c*.

Before we added $\Diamond$, our toy dynamic semantics contained only persistent sentences.

It should be evident that this semantics leads to a failure of distributivity. Whether a set of worlds as a whole contains a ϕ-world is a 'global' property of that set. Thus if one wants to know whether some $\neg\phi$-world w left open by a context c will survive an update

with $\Diamond\phi$, the answer will not be settled by some property of w; rather, one needs to know something about what other worlds there are in c.

All this makes it especially confused to speak of the "content" or "proposition expressed" by $\Diamond\phi$ in a dynamic setting. Such sentences do not semantically serve to represent the world as being a certain way, or serve to add a proposition to the common ground; their discourse impact is just different from that of straight factual sentences. Nor is it helpful to say that $\Diamond$ corresponds to a special "force" on this account. Speech act forces, whatever those are, don't embed, at least on the usual way of understanding them; but epistemic modals are freely embeddable on the dynamic account.

The point that epistemic modal claims don't determine truth-conditions—that they aren't the kind of clauses that determine conditions on points, the sets of which constitute contexts—calls into question the assumption, not uncommon in discussions of presupposition projection, that we can always factor out a truth-conditional contribution for any dynamic connective.

What is the empirical motivation for this dynamic semantics? Perhaps the most frequently cited data point in its favor concerns an alleged asymmetry in acceptability between the following two sorts of discourse:

(26) Billy might be at the door (and) . . . it isn't Billy at the door. ($\Diamond\phi \wedge \neg\phi$)
(27) ? It isn't Billy at the door (and) . . . Billy might be at the door. ($\neg\phi \wedge \Diamond\phi$)

—Where we are to imagine that what happens between the dots is that the speaker sees who is at the door. Pairs like this one were first discussed by Veltman, and are cited approvingly by Groenendijk et al. [1995], Veltman [1996], Beaver [2001], Gillies [2001], and von Fintel and Gillies [2007], among others. The dynamic account predicts the supposed difference between these discourses. The update corresponding to (27) is guaranteed by dynamic lights to "crash" the context, since for all c and ϕ, $c[\neg\phi \wedge \Diamond\phi] = c[\neg\phi][\Diamond\phi]$, which in turn will always be $\emptyset$. Once the context comes to accept $\neg\phi$, it will fail the test corresponding to $\Diamond\phi$. But not so for (26), the other ordering of the conjuncts: $c[\Diamond\phi][\neg\phi]$ will be nonempty as long as c contains some ϕ-worlds and some $\neg\phi$-worlds. The thought, then, is that pairs like (26)–(27) demonstrate the need for *both* a dynamic semantics for epistemic possibility modals *and* a dynamic semantics for conjunction, for these together generate the order-sensitivity (noncommutivity) manifested by such pairs.

If correct, this semantics would, as Beaver suggests, help to constrain the space of possible context change potentials for connectives *vis-à-vis* the problem of presupposition projection. For example, on this semantics it is no longer the case that the ordinary conjunction (C1) and deviant conjunction (C2) have the same update impact where both defined. On the contrary, they, would come apart exactly on pairs such as (26)–(27). If ordinary conjunction were defined as in (C2), then (27) would be predicted to be felicitous and (26) infelicitous—the incorrect result, says the fan of dynamics for epistemic modals.

But, stepping back, what is the right thing to say about pairs such as (26)–(27)?

First, it should be duly noted that sentences of the form ($\Diamond\phi \wedge\neg\phi$) and ($\neg\phi \wedge\Diamond\phi$) are virtually always both marked, and equally so. Generally they sound just as bad as (e.g.) contradictions. Second: even if, with sufficient rigging of context, (26) can be made to sound okay, this alone tells us nothing; for with sufficient rigging, (27) will sound

okay, as will almost anything else. What then is the point? The point is supposed to be that pairs such as (26)–(27) manifest an asymmetry in markedness in a certain kind of situation: namely, one in which *monotonic* information growth happens in context mid-sentence. The thought is that if we restrict to such cases, we will see that (26), but not (27), will sound fine. The is supposed to highlight the need for a nonpersistent update semantics for $\Diamond\phi$ with noncommutative conjunction.

Alas, it does not. First, the data do not support noncommutivity for conjunction. The sentence (26) is marginal at best with *and*, even in the scenario described. One senses some awareness of this fact in the literature, for among the works cited above, only one (namely Beaver [2001]) actually explicitly includes *and* in the relevant pairs. Beaver is of course correct to include it—to draw conclusions about the semantics of an expression, one presumably should appeal to constructions in which the expression actually occurs—but judgments do not support non-commutativity.

Moreover, if things are as the dynamic story is telling us, one wonders why there is no substantial asymmetry in defectiveness between

(28) ?? Suppose Billy might be at the door and he isn't.
(29) ?? Suppose Billy isn't at the door and he might be.

According to the dynamic account as so far stated, (29) is predicted to be defective, but not (28). But the two are both defective, and intuitively, they are defective in the same way. Perhaps there is some further story the fan of dynamics will tell here, but until then, we have no data supporting non-commutative conjunction. If anything, we have data favoring commutativity.

Second, the way that the notion of monotonic information growth is used in setting up the data is questionable. To see why, consider a view according to which the transition from believing $\Diamond\phi$ to believing $\neg\phi$ normally involves nonmonotonic information growth. Egan [2007] is perhaps one such example (see also Stephenson [2007]): on his view, roughly, to believe that Billy might be at the door is to "locate oneself", in the sense of Lewis [1979], on a person whose evidence leaves the possibility of Billy's being at the door open, where the possibilities left open by one's evidence are centered worlds. In the background scenario stipulated for (26)–(27), the speaker's belief state would, on Egan's account, normally nonmonotonically change in respect of the speaker's views about his evidence: his belief state would go from excluding centered worlds where the center's evidence entails Billy is not at the door to including such worlds. Now we can ask: Is there reason for Egan to be phased by anything about (26)–(27)? It seems not: he can simply deny that there really is strictly monotonic information growth in the (26)-case, and just explain that insofar as the sentences are felicitous, it is because they express propositions rationally believed at the relevant times of utterance. Could it be objected against Egan that it is intuitively obvious that the (26)-case involves only monotonic information growth? No, for it makes little sense to appeal to brute intuitions about the technical notion of monotonic information growth.

We conclude pairs such as (26)–(27) provide no evidence for an especially dynamic account of epistemic possibility modals, or of conjunction.

But what of pairs such as (28)–(29)? Although they do not support a non-commutative semantics for conjunction, they *do* support the idea that there is some semantically significant kind of incompatibility between $\Diamond\phi$ and $\neg\phi$. The defectiveness of examples such as (28) and (29) is in fact hard to predict on the leading static

semantics for epistemic modals, namely that of Kratzer [1991]; similiarly with the accounts of Egan [2007] and Stephenson [2007]. On these accounts $\Diamond\phi$ and $\neg\phi$ are truth-conditionally compatible, and so one expects it to be possible to suppose them jointly true. But on a dynamic conception the sentences are incompatible, inasmuch as they are not jointly acceptable by nonempty contexts; as a result, on usual dynamic ways of defining consequence (such as the definition considered in 2.2 above), it will come out that $\neg\phi \models \neg\Diamond\phi$. Unsurprisingly then, when the dynamic account is suitably upgraded with semantics for attitude verbs (e.g., Heim [1992]; see also Asher [1987]), it gets the right results for (28) and (29), and for related constructions. And this is a indeed a real advantage over its static rivals.

Though not all of them. The static account of Yalcin [2007] can accommodate, indeed is motivated by, these data (see also Kolodny and MacFarlane [2010], MacFarlane [2011]). The account of that paper does not assume DYNAMIC, but it does define semantic notions of acceptance and of consequence entirely analogous to those usual in a dynamic semantics, and it gets the correct results for sentences like (28) and (29). The abstract idea behind Veltman's semantics, that the transition from believing or presupposing $\Diamond\phi$ to ceasing to believe or presuppose $\Diamond\phi$ is generally monotonic, certainly has empirical motivation. (Though see Willer [2010a], Yalcin [2011a] for some further problems and extensions.) But it can be realized in a static framework.

6 Essentially Dynamic?

Because of the way in which Veltman's account of epistemic modals introduces violations of distributivity and persistence, it is sometimes suggested that his modals are "essentially dynamic operators" (Muskens et al. [2011]; cf. von Fintel and Gillies [2007]). But inlight of our recent observations, this assessment seems somewhat hasty. It is true, modulo a certain kind of radical context-sensitivity, that one cannot resolve Veltman's proposed context change potentials into static conditions on points—truth-conditions, in the usual sense—together with a simple intersective rule of context update. But one can effectively resolve it into a static condition on bodies of information together with a single rule mapping such static conditions and prior contexts into posterior contexts. And if one can do this, why regard the underlying idea as essentially dynamic? (Of course, a dynamic formulation might prove to be more elegant in various respects that its static counterpart, hence to be preferred for that reason; but that is another story.) For further discussion of this subtle issue, see Rothschild and Yalcin [unpublished].

7 A Few More References

For more on presupposition and anaphora see, in addition to the citations already given, Van der Sandt [1992], Geurts [1999], Breheny [2001], Rothschild [2008a], Kripke [2009]. For a start on dynamics for temporal anaphora, see Ter Meulen [1997], [Steedman, 1997, sec. 3.2.3], Stone [1997], Stone and Hardt [1999]. For a start on dynamics for modal anaphora, see Roberts [1989, 1996], Frank and Kamp [1997], Stone [1997], Brasoveanu [2007].

Conditionals form an active area of dynamic semantic research. On counterfactuals, see von Fintel [2001], Gillies [2007], Moss [2010]. See Gillies [2004] for an account of indicative conditionals in the spirit of Veltman's update semantics for epistemic modals.

See also Willer (2011b) See Yakin for a dynamic account of indicative conditionals incorporating probabilities.

Interrogatives and focus have formed another active area of dynamic semantic inquiry. See Ginzburg [1996], Erteschik-Shir [1997], Hulstijn [1997], Groenendijk [1999], Aloni and Van Rooy [2002], and the papers in Aloni et al. [2007].

References

Maria Aloni and Robert Van Rooy. The dynamics of questions and focus. In *Proceedings of SALT*, volume 12, pages 20–39. Citeseer, 2002.

Maria Aloni, Alastair Butler, and Paul Dekker, editors. *Questions in Dynamic Semantics*, volume 17 of *Current Research in the Semantics/Pragmatics Interface*. Emerald Group Publishing, 2007.

Nicholas Asher. A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy*, 10(2): 125–197, 1987.

David Beaver. *Presupposition and Assertion in Dynamic Semantics*. Studies in Logic, Language and Information. CSLI Publications, 2001.

David Beaver. As brief as possible (but no briefer). *Theoretical Linguistics*, 34(3): 213–228, 2008.

Adrian Brasoveanu. *Structured Nominal and Modal Reference*. PhD thesis, Rutgers, New Brunswick, NJ, 2007.

Richard Breheny. Indefinites and anaphoric dependence: A case for dynamic semantics or pragmatics? In M. Reimer and A. Bezuidenhout, editors, *Descriptions and Beyond*. Oxford University Press, Oxford, 2001.

Robin Cooper. The interpretation of pronouns. *Syntax and Semantics*, 10: 61–92, 1979.

Sam Cumming. Variabilism. *The Philosophical Review*, 117(4): 525, 2008.

Andy Egan. Epistemic modals, relativism and assertion. *Philosophical Studies*, 133 (1): 1–22, 2007.

Paul D. Elbourne. *Situations and Individuals*. MIT Press, Cambridge, MA, 2005.

Paul D. Elbourne. The interpretation of pronouns. *Language and Linguistics Compass*, 2(1): 119–150, 2008.

Nomi Erteschik-Shir. *The Dynamics of Focus Structure*. Cambridge University Press, Cambridge, 1997.

Gareth Evans. Pronouns. *Linguistic Inquiry*, 11(2): 337–362, 1980.

Kai von Fintel. Counterfactuals in a dynamic context. In Ken Kenstowicz, editor, *Ken Hale: A Life in Language*, pages 123–152. MIT Press, Cambridge, MA, 2001.

Kai von Fintel and Anthony Gillies. An opinionated guide to epistemic modality. *Oxford Studies in Epistemology*, 2, 2007.

Danny Fox. Two short notes on Schlenker's theory of presupposition projection. *Theoretical Linguistics*, 34(3): 237–252, 2008.

Anette Frank and Hans Kamp. On context dependence in modal constructions. In *Proceedings of SALT VII*, pages 21–23, 1997.

Gerald Gazdar. *Pragmatics: implicature, presupposition and logical form*. Academic Press, 1979.

P.T. Geach. *Reference and Generality*. Cornell University Press, Ithaca, NY, 1962.

Bart Geurts. *Presuppositions and pronouns*. Elsevier, 1999.

Bart Geurts and David Beaver. Discourse representation theory. *Stanford Encyclopedia of Philosophy*, 2007.

Anthony Gillies. A new solution to Moore's paradox. *Philosophical Studies*, 105: 237–250, 2001.

Anthony Gillies. Epistemic conditionals and conditional epistemics. *Noûs*, 38: 585–616, 2004.

Anthony Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3): 329–360, 2007.

Jonathan Ginzburg. Interrogatives: Questions, facts and dialogue. *The Handbook of Contemporary Semantic Theory*, 5:359–423, 1996.

Jeroen Groenendijk. The logic of interrogation. In *Proceedings of SALT IX*. Institute for Logic, Language and Computation (ILLC), University of Amsterdam, 1999.

Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.

Jeroen Groenendijk and Martin Stokhof. Modality and conversational information. *Theoretical Linguisitics*, 2:61–112, 1975.

Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and modality. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell, 1995.

Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, 1982.

Irene Heim. File change semantics and the familiarity theory of definiteness. In R. Bauerle, C. Schwarze,

and A. Von Stechow, editors, *Meaning, Use and Interpretation of Language*, pages 164–90. De Gruyter, 1983a.

Irene Heim. On the projection problem for presuppositions. In M. Barlow, D. Flickinger, and M. Wescoat, editors, *WCCFL 2: Second Annual West Cost Conference on Formal Linguistics*, pages 114–25. Stanford University, 1983b.

Irene Heim. Presupposition projection. In *Reader for the Nijmegen Workshop on Presupposition, Lexical Meaning, and Discourse Processes, Univ. of Nijmegen*, 1990a.

Irene Heim. E-type pronouns and donkey anaphora. *Linguistics and philosophy*, 13 (2): 137–177, 1990b.

Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9: 183–221, 1992.

Joris Hulstijn. Structured information states: raising and resolving issues. In A. Benz and G. Jager, editors, *Proceedings of MunDial97*, University of Munich, 1997.

Pauline Jacobson. Towards a variable-free semantics. *Linguistics and Philosophy*, 22 (2): 117–185, 1999.

Hans Kamp. A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof, editors, *Formal Methods in the Study of Language*. Mathematisch Centrum, University of Amsterdam, Amsterdam, 1981.

Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Springer, 1993.

Lauri Karttunen. What makes definite noun phrases definite? Technical Report P3871, RAND Corporation, Santa Monica, CA, 1968.

Lauri Karttunen. Discourse referents. In J. McCawley, editor, *Syntax and Semantics 7*, pages 363–85, Academic Press, New York, 1976.

Lauri Karttunen and Stanley Peters. Conventional implicature. In C. K. Oh and D. Dineen, editors, *Presupposition*, Syntax and Semantics vol. 11, pages 1–56. Academic Press, 1979.

Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, CVII(3): 115–43, March 2010.

Angelika Kratzer. Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*, pages 639–50. De Gruyter, 1991.

Saul A. Kripke. Presupposition and anaphora: Remarks on the formulation of the projection problem. *Linguistic Inquiry*, 40(3): 367–386, 2009.

David K. Lewis. General semantics. *Synthese*, 22:18–67, 1970.

David K. Lewis. Adverbs of quantification. In E. L. Keenan, editor, *Formal Semantics of Natural Language*, pages 3–15. Cambridge University Press, Cambridge, 1975.

David K. Lewis. Attitudes de dicto and de se. *Philosophical Review*, 88(4): 513–43, 1979.

John MacFarlane. Epistemic modals are assessment-sensitive. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*. Oxford University Press, Oxford, 2011.

Sarah Moss. On the pragmatics of counterfactuals. *Noûs*, 2010.

Reinhard Muskens. Combining Montague semantics and discourse representation. *Linguistics and Philosophy*, 19(2):143–186, 1996.

Reinhard Muskens, Johan van Benthem, and Albert Visser. Dynamics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 607–670. Elsevier, 2nd edition, 2011.

Dilip Ninan. *Imagination, Content, and the Self*. PhD thesis, MIT, 2008.

Dilip Ninan. De se attitudes: Ascription and communication. *Philosophy Compass*, 5 (7): 551–567, 2010.

Terry Parsons. Pronouns as paraphrases. Unpublished manuscript, University of Massachusetts, Amherst, 1978.

Craige Roberts. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy*, 12(6): 683–721, 1989.

Craige Roberts. Anaphora in intensional contexts. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell, Oxford, 1996.

Daniel Rothschild. Presupposition projection and logical equivalence. *Philosophical Perspectives*, 22(1): 473–497, 2008a.

Daniel Rothschild. Transparency Theory and its dynamic alternatives: Commentary on "Be Articulate". *Theoretical Linguistics*, 34(3): 261–268, 2008b.

Daniel Rothschild. Explaining presupposition projection with dynamic semantics. Semantics and Pragmatics, 4(3): 1–43, 2011.

Daniel Rothschild and Seth Yakin. Dynamics. Unpublished manuscript, 2011.

Philippe Schlenker. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, 34(3): 157–212, 2008.

Philippe Schlenker. Local contexts. *Semantics and Pragmatics*, 2(4): 1–78, 2009.
Scott Soames. Presupposition. *Handbook of Philosophical Logic*, 4: 553–616, 1989.
Robert Stalnaker. Pragmatics. *Synthese*, 22: 272–89, 1970.
Robert Stalnaker. Pragmatic presuppositions. In M. K. Munitz and P. Unger, editors, *Semantics and Philosophy*. NYU Press, New York, 1974.
Robert Stalnaker. Indicative conditionals. *Philosophia*, 5, 1975.
Robert Stalnaker. Assertion. In P. Cole, editor, *Syntax and Semantics 9: Pragmatics*. Academic Press, 1978.
Robert Stalnaker. *Inquiry*. MIT Press, Cambridge, MA, 1984.
Robert Stalnaker. On the representation of context. *Journal of Logic, Language, and Information*, 7: 3–19, 1998.
Mark Steedman. Temporality. In Alice ter Meulen and Johan van Benthem, editors, *Handbook of Logic and Language*, pages 895–938. Elsevier, 1997.
Tamina C. Stephenson. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30(4): 487–525, 2007.
Matthew Stone. The anaphoric parallel between modality and tense. Technical Report 06, Institute for Research in Cognitive Science, University of Pennsylvania, 1997.
Matthew Stone and Daniel Hardt. Dynamic discourse referents for tense and modals. *Computing Meaning*, 1: 301–320, 1999.
Alice Ter Meulen. *Representing Time in Natural Language–The Dynamic Interpretation of Tense and Aspect*. MIT Press, Cambridge, MA, 1997.
Johan van Benthem. *Essays in Logical Semantics*. Studies in Linguistics and Philosophy. Reidel, Dordrecht, 1986.
Johan van Benthem. *Exploring Logical Dynamics*. Center for the Study of Language and Information, Stanford, CA, 1996.
Robert A. Van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4): 333, 1992.
Jan van Eijck and Hans Kamp. Discourse representation in context. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 181–252. Elsevier, 2nd edition, 2011.
Frank Veltman. *Logics for Conditionals*. PhD thesis, University of Amsterdam, Amsterdam, 1985.
Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3): 221–61, 1996.
Malte Willer. *Modality in Flux*. PhD thesis, University of Texas, 2010a.
Malte Willer. New surprises for the Romsey Test. *Synthese*, 176(2): 291–309, 2010b.
Seth Yalcin. Epistemic modals. *Mind*, 116(464): 983–1026, Nov. 2007.
Seth Yalcin. Nonfactualism about epistemic modality. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*. Oxford University Press, Oxford, 2011a.
Seth Yalcin. A counterexample to Modus Tollens. *Journal of Philosophical Logic*, 2011b.

2.8

EVENT SEMANTICS

Barry Schein

Davidson's diamonds are a recurrent pattern of inference in natural language, in which elimination of a modifier phrase or argument, the arrows in (1), is valid; but, introduction reversing the arrows is not (Davidson 1967, 1985; Castañeda 1967; Parsons 1985, 1990):

(1) Jones (slowly) buttered the toast with a knife in the kitchen.
↙↘
Jones (slowly) buttered the toast with a knife. Jones (slowly) buttered the toast in the kitchen.
↘↙
Jones (slowly) buttered the toast.
↙↘
Jones (slowly) buttered. The toast was (slowly) buttered.
↘↙
There was a (slow) buttering.

(2) $\exists e(\text{Agent}(e, x_1) \,\&\, \text{slow}(e) \,\&\, \text{butter}(e) \,\&\, \text{Patient}(e, x_2) \,\&\, \text{with}\,(e, x_3) \,\&\, \text{in}(e, x_4))$

A Davidsonian logical form (2), being a conjunction of terms, explains that elimination of a modifier or argument is elimination of a conjunct. It further represents that all the conjuncts are about a particular *e* in a sentence which asserts that an *e* exists to satisfy them all. Introduction is thus shown to be a mistaken inference from the existence of some particulars satisfying given descriptions to the existence of a single one satisfying a conjunction of them:

(3) $\exists e(\text{Agent}(e, \text{j}) \,\&\, \text{butter}(e)\ [\text{an } x_3\text{: knife}(x_3)]\ \text{with}(e, x_3))$
Jones buttered with a knife.
$\exists e(\text{Agent}(e, \text{j}) \,\&\, \text{butter}(e) \,\&\, [\text{the } x_4\text{: kitchen}(x_4)]\ \text{in}(e, x_4))$
Jones buttered in the kitchen.
$\nvdash \exists e(\text{Agent}(e, \text{j}) \,\&\, \text{butter}(e) \,\&\, [\text{an } x_3\text{: knife}(x_3)]\ \text{with}(e, x_3) \,\&\, [\text{the } x_4\text{: kitchen}(x_4)]\ \text{in}(e, x_4))$
Jones buttered with a knife in the kitchen.

Buttering with a knife in the pantry and again in the kitchen with a spatula does not imply that Jones ever took knife to butter in the kitchen.

In the logical forms for (1), the morpheme *butter* emerges with a single meaning for all its occurrences in various syntactic and morphological contexts. Variable polyadicity ranging from zero to four arguments is just an illusion of the monadic *butter* in construction with a varying number of conjuncts. Logical form here answers to both inference in natural language and the univocality of its morphemes. Event semantics, so called, comprises those approaches to inference and grammar that would, in agreement with (2), introduce new parameters and decompose the simple sentence into a conjunction of terms containing them. Events, by the way, are the values for *e* in true sentences. [See Schein 2002 for more on the metaphysics of events. Speakers affirm (i)-(iii) without implying (iv) or (v), a dilemma for Davidsonian event semantics as its uncomplicated logical forms for (i)–(iii) do imply (iv) and (v) (see Lombard 1985).

(i) The Carnegie Deli faces Carnegie Hall.
(ii) Carnegie Hall faces the Carnegie Deli.
(iii) The Carnegie Deli facing Carnegie Hall *is* Carnegie Hall facing the Carnegie Deli.
(iv) F The Carnegie Deli faces the Carnegie Deli.
(v) F Carnegie Hall faces Carnegie Hall.

A necessary complication is to recognize that the material states identified in (iii) are not the observations under a perspective or scenes of those states that distinguish (i) and (ii) and undo the unwanted implication. The scenes or perspectives that here rescue event semantics from itself are however pervasive wherever there is predication.]

This chapter rummages for more gems of natural language to fix the logical syntax of talk about events in simple sentences. The existential event quantification with scope over the entire sentence in (2) turns out to be plural, and the cross-reference to events from one conjunct to another is by plural pronoun or definite description (*plural event quantifiers and pronouns*) rather than bare variable. Instead of relating to the same event *e* as in (2), every argument prompts its own events in a more extensive decomposition (*supermonadicity*), in which various relations between the several events hold the sentence together (see (25) below). Emending how quantifier phrases and descriptions prefix to their scope, *adverbialization* interposes an adverb derived from the nominal phrase's descriptive content so that what is spoken as (4), for example, is always parsed along the lines of (5):

(4) The butler buttered.
(5) The butler *while a butler* buttered.

These emendations retain the original explanation for the validity of modifier and argument elimination and for the invalidity of their introduction; but, the logical form that replaces (2) paraphrased sounds more like (6):

(6) Jones while Jones participated & it was slow & it caused a buttering & the toast while toast participated in that, which had assistance which a knife while a knife participated in, & it was a being-in which the kitchen while a kitchen participated in.

It gets worse with annotations in logical form referring to perspectival events or states, scenes or frames of reference for the outside events the sentence is a report of. Event

semantics, its syntax revised and the architecture of simple clauses fortified, is then found to resolve problems of extensional substitutivity under identity and to clarify the logical form of identity statements in natural language, deploying event talk to safeguard naive reference into an austere ontology.

§1.1. *Plural event pronouns* make a quick entrance behind adverbs denoting events. Contrary adverbs (e.g., *gracefully vs. clumsily*) prove the source of syntactical insight:

(7) Jones gracefully and Godfrey clumsily buttered 613 pastries for brunch.

As what is graceful is not clumsy, Jones buttering 365 brioches and Godfrey, 248 croissants, attend to brunch with their own events. Their graceful and clumsy actions together butter the pastries. Saying so requires two more existential event quantifiers, and a plural pronoun dividing its reference between the two events supplants a simple event variable in relating action to buttering:

(8) $\exists e\, \exists e_1\, \exists e_2$((Agent($e_1$, j) & graceful($e_1$)) and (Agent($e_2$, g) & clumsy($e_2$)) & [$\wr E$: Ee_1 & Ee_2] butter 613 pastries[e, E])

[Atomic formulae, e.g., 'butter(e)', 'Ee', appear with variables bare or in parentheses. Brackets indicate a molecular formula with the variables enclosed free, e.g., 'butter the toast[e,E]'. N.B. In 'Agent[E,X]', 'cluster[E]', *etc*, the brackets flag abbreviation of complex, second-order formulas composed of first-order relations—not primitive, higher-order relations, *v*. (15), (18).]

Note that *and* is univocally the sentential connective. The burden of collective predication falls on the plural event pronoun here and also in the logical form for (9), identical to (8) except for omission of the adverbs.

(9) Jones and Godfrey buttered 613 pastries.

Contrary adverbs recur in (10) and (11) with similar mission and further instruction:

(10) Jones gracefully and Godfrey clumsily separated the brioches to the left and the croissants to the right.
(11) At the loom, Arachne gracefully and her apprentice clumsily criss-crossed silk threads horizontally and linen threads vertically.

Neither the leftward nor the rightward parade of pastry is a separating on its own; and, the separating is itself neither to the left nor right, having components of both. So neither the leftward event nor the rightward event is described by the verb, and the adverbs modify a separate relation, Theme in (12), describing the brioches' and the croissants' participation (*pace* Kratzer 1996, 2003):

(12) $\exists e\, \exists e_1 \exists e_2 \exists e_3 \exists e_4$ ((Agent(e_1,j) & graceful(e_1)) and (Agent(e_2,g) & clumsy(e_2)) & [$\wr E_{12}$: $E_{12}e_1$ & $E_{12}e_2$] [$\wr E_{34}$: $E_{34}e_3$ & $E_{34}e_4$] separate[e, E_{12}, E_{34}] & (Theme(e_3,b) & to the left[e_3]) and (Theme(e_4,c) & to the right[e_4]))

Another plural pronoun referring to the leftward and rightward events must, in turn, relate the brioches' and croisssants' participation to the separating, an e that separate(e).

To continue in the spirit of (2) until it is abandoned below for supermonadicity, (10) could still be held to be about a single separating and the graceful and clumsy on the one hand and the leftward and rightward on the other could just be two ways to parse it into mereologically smaller events, into those where Jones and Godfrey are Agents and into those where the brioches and the croissants are Themes.

Plural event pronouns have been introduced to collect the events in which participants participate, and they divide their reference accordingly. Verbs also denote plural events and so then must the quantifier prefixed to the sentence. *Plural existential event quantification* (Gillon 1990; Pietroski 2005; Schein 1993, 2006; Schwarzschild 1991, 1996) with a verb denoting plural events is found in (13), given that the same morpheme *cluster* occurs twice:

(13) The fires will cluster in two clusters (and not in one).
(14) $\exists E$ ([The X: fires(X)] Theme(E,X) & cluster[E] . . .))
(15) cluster[E] $\leftrightarrow_{\text{def}} \exists e Ee$ & $\forall \exists e(Ee \rightarrow \text{cluster}(e))$

Divine law and faith that the events scattered in history are distinct find plural event quantification in (16) too:

(16) In the era chronicled, two Roman bishops against their will and two Anglican churchmen at the behest of the throne have joined three English kings eagerly and ten French princesses reluctantly in holy matrimony.

Here the several joinings the verb denotes are equal in number to the sum of events executed against the will and events at the behest of the throne and fewer than the events submitted to eagerly and reluctantly; but, they are all in the plural.

§1.2. A more extensive decomposition of polyadicity is next, *supermonadicity*. In (16), the clerics are all joiners and so all Agents. In the conjunction of predicate phrases in (17), the first describes provocateurs, Agents in actions at the Pentagon and the second, victims, that is, Patients, acted upon on the Mall:

(17) The Columbia students (noisily) and the Harvard students (quietly) surrounded the Pentagon and were crowded into the Mall.
(18) $O[E_i, E_j] \leftrightarrow_{\text{def}} \exists \forall e(\exists e_i(E_i e_i$ & overlap(e,e_i)) $\leftrightarrow \exists e_j(E_j e_j$ & overlap(e,e_j)))
(19) $\exists e\ \exists e_1 \exists e_2 \exists e_3 \exists e_4$(Participate($e_1$, C) and Participate($e_2$, H) &
$[\neg E_{12}\colon E_{12}e_1$ & $E_{12}e_2][\neg E_{34}\colon E_{34}e_3$ & $E_{34}e_4]$ O[E_{12}, E_{34}] &
$\exists e_3(\exists X$ Agent(e_3, X) & surround the Pentagon[e_3,X]) and
$\exists e_4(\exists X$ Patient(e_4, X) & be crowded into the Mall[e_4, X])

The Columbia students participated, and the Harvard students participated; & it all was a surrounding the Pentagon and a being crowded into the Mall.

Sentence (17) commits a student to neither venue, leaving vague whether she acted at the Pentagon or was acted upon on the Mall. All that is implied is that the Columbia students participate in some way as do the Harvard students, and what they do constitutes the events subsequently described, so that some of the Columbia and Harvard students must have been at the Pentagon while the rest were on the Mall. The conjuncts in (19) mentioning the students are necessarily vague about what they are up to in

their noisy and quiet events, asserting only their participation. That vagueness however threatens translation of (20). In the event that Harvard students mercilessly tease Columbia students and Columbia students decline to tease themselves, (20) is false:

(20) The Harvard students and the Columbia students teased the Columbia students.
(21) The Harvard students participate and the Columbia students participate; & it was a teasing.

But its considered translation paraphrased in (21) is mistakenly true in that both the Harvard students and the Columbia students have participated, the former as aggressors and the latter as victims in the same event of teasing. To divorce the vagueness necessary for (17) from its ill effect on (20), the logical form for the latter must resolve two events, cause and effect:

(22) $\exists e_1 \exists e_2 \exists e_3$(Participate($e_1$, C) and Participate($e_2$, H) & $[\imath E_{12}\colon E_{12}e_1$ & $E_{12}e_2][\imath E_3\colon E_3e_3]$ Cause(E_{12}, E_{34}) & tease(e_3) & Participate(e_3, C))

The Columbia students participate and the Harvard students participate; & it caused a teasedness (cf. 'embarrass', 'shame', 'humiliate'); *& the Columbia students participate in that*.

The vagueness of the Harvard and Columbia students' participation notwithstanding, what they do is said to cause the Columbia students' humiliation. The Columbia students had no hand in the causal event, and (22) is false, as a translation of (20) should be.

This argument repeats itself *mutatis mutandis* wherever nominal phrases can occur allegedly conjoined, and thus a resolution into smaller events occurs across-the-board at every argument position. [E.g., *613 diverse hedge funds variously sold and sold to 365 equity funds and 248 mutual funds in a single day's trading. The preparation of these Balkan delicacies has variously stuffed full and stuffed into other vegetables— thirty eggplants, fifty peppers, eighty tomatoes, 130 olives and 210 pearl onions*.] [Pietroski (1998), arguing from rather different considerations, also reaches the conclusion that thematic relations themselves occasion the introduction of an event quantifier. A dyadic relation 'C(e_1, e_2)', if e_1 is meant to be a cause and e_2 its effect, leaves no room for reference to the event that is cause and effect combined, for which purpose one may prefer to substitute a triadic relation 'C(e_1,e_2,e_3)'. See Pietroski 1998, 2000 and references cited and Schein 2002.]

In the initial Davidsonian logical form (24) for (23), relations to events relate to the same event *e*:

(23) Jones buttered the toast in the kitchen.
(24) Agent(e, x_1) & butter(e) & Patient(e, x_2) & in(e, x_3)
(25) Participate(e_1, x_1) & Cause(e_1,e_2) & butter(e_2) & Participate(e_2, x_2) & in(e_2, e_3) & Participate(e_3, x_3)

With *supermonadicity*, it is small events all the way down, with every nominal phrase, including each conjoined (see (11)), launching its own. It doesn't get lonelier than

that, spinning around in a subevent solo, unrelated to anything else except through the mediation of further relations between events. This revision to clausal architecture is warranted by three elementary observations: i. sentences with multiple conjunctions (17) demand a certain vagueness about the manner of participation in the events described; ii. this vagueness should not then confuse aggressors and victims (20); and iii. any participant's participation may be described by an adverb that is contrary to how anyone else participates ((7), (10), (11), (17)).

§1.3. *Adverbialization* of nominal descriptive content is a fact of grammar. First, descriptive anaphora, both implicit and explicit as in (26), co-varying with the events described, must fall within the scope of the event quantifier adverbialization derives as in (28):

(26) Every fugitive from a federal prison is soon apprehended near it. (after Enç 1986)
(27) Every fugitive from a federal prison$_j$ *as of* ***some***$_i$ ***time*** is soon *after* ***that***$_i$ ***time*** apprehended near it$_j$.
(28) [Every *x*: ∃*e* fugitive from a federal prison[*e*, *x*]]
[*while e*: fugitive from a federal prison [*e*, *x*]]
∃*e*'(soon(*e*, *e*') & apprehended[*e*', *x*] &
[ι *y*: federal prison(*y*) from which *x* is a fugitive in *e*] near(*e*', *y*))

Second, the adjective *respective* can only be interpreted as in some way modifying the matrix scope of its host nominal phrase, as if it were the adverb *respectively* (Gawron & Kehler 2002, 2004). Suppose that twins rent the same make and model vehicle and pose for photographs with the two cars in which it is variously true that:

(29) The twins stood beside their respective rentals.
The twins stood behind their respective rentals.
(30) The twins stood between their rentals.

In contrast to (30), there can be no photograph for which:

(31) *The twins stood between their respective rentals.

For the twins to be in respective possession of their rentals, as the nominal describes them, puts them in a pair of states in each of which a twin is opposite her own rental. Such a state frames a standing beside (or a standing behind) if its twin is beside (or behind) her rental, (29). But, neither of the states that respective possession denotes frames a standing of a twin between the rental in it, (31). This nonsense that (31) cannot escape is a consequence of adverbialization demanding that the twins being with their respective rentals frame their standings between.

The scope of adverbialization tracks the scope of the nominal phrase adverbialized:

(32) At the thriller in Manila, Muhammad Ali's trainer and Joe Frazier's trainer pulled their respective champions apart.
(33) At the thriller in Manila, Muhammad Ali's trainer and Joe Frazier's trainer pulled apart their respective champions.

Given the scope of *their respective champions* in (32), it is immediate that the states of respective possession separately frame only the champions being pulled: Ali's trainer grabbing Ali and Frazier's trainer grabbing Frazier pull hard until the champions are parted. Later noticed is the macabre interpretation in which the states of respective possession each frame a champion being pulled apart by his trainer. The macabre becomes more salient in (33) to the extent that stylistic movements favor the scope worn on the sleeve, here the one now including *apart* within the scope of *their respective champions* (slighted is the variation from speakers like myself for whom (33) is unambiguously macabre to those for whom the contrast between (32) and (33) is rather weak.). Note that for the only sensible interpretation of (32) and (33), the scope that adverbialization affords to *respective* is the minimal one that the host nominal phrase itself requires, a relation expressing participation in some events—a scope that is too small for the felicitous use of the proper lexical adverb *respectively*:

(34) * At the thriller in Manila, Muhammad Ali's trainer and Joe Frazier's trainer respectively pulled their champions apart.
(35) * At the thriller in Manila, Muhammad Ali's trainer and Joe Frazier's trainer pulled apart their champions respectively.
(36) ?? At the thriller in Manila, Muhammad Ali's trainer and Joe Frazier's trainer pulled their champions respectively— apart.

Nominal descriptions are the adverbs dedicated to the smallest relations to events, reflecting an economy of expresssion and compactness in natural language, which reserves lexical adverbs for larger phrases containing whole verbs.

['While' appears in the translations above for illustration. Expected is the range of relations, temporal, causal, causal explanatory, mediating the interpretation of *if*-clauses or absolutive clauses including variation therein due to differences between tensed and reduced clauses.

Adverbialization is grammar sometimes without apparent effect framing the events reported, *The liar was deep asleep*—but not while lying. Compare:

(i) The liar fails a polygraph.
Any liar fails a polygraph.
Every liar fails a polygraph.

(ii) The sometimes liar fails a polygraph.
Any sometimes liar fails a polygraph.
Every sometimes liar fails a polygraph.

As expected from adverbialization, the polygraph failures of (i) are an effect of lying, and any lie that escapes detection is a counterexample to what is asserted. In (ii), it seems that sometimes lying makes one anxious enough to fail a polygraph even when telling the truth. To be a sometimes liar is to *always* be a sometimes liar, for as long as one lives. So, to do something while a *sometimes* liar is to do something while oneself. Thus, *The liar was deep asleep* is *The sometimes liar while a sometimes liar was deep asleep*, with *sometimes* unspoken.]

§2. *Extensional Substitutivity under Identity*. (Barber 2000; Carlson 1982, Forbes 1997, 1999; Predelli 1999; Saul 1997 and, 1999, 2007; Szabó 2003)

Adverbialization (see Forbes 1999) joins forces with supermonadicity to resolve puzzles of extensional substitutivity under identity:

(37) The ponds the temperatures of which were recorded early this morning were colder than the ponds the temperatures of which were recorded early this evening.

(38) Every pond the temperature of which was recorded early this morning was colder than every pond the temperature of which was recorded early this evening.
(39) Venus Pond early this morning was colder than Venus Pond early this evening.
(40) * Venus Pond was colder than Venus Pond (was).

From neither (37) nor (38) is there implication that the morning ponds are other than the evening ponds. If, in fact, Venus Pond (Maui) is among them both, (37) and (38) imply (39). Insofar as adverbial qualification is necessary to escape contradiction (cf. (40)), its only source in (37) and (38) are the nominal phrases themselves which, according to adverbialization, do indeed supply adverbial phrases to qualify participation in the events described.

As morning is not evening and manhood not boyhood, the adverbs from adverbialization in (37) and (38) and in (41) cannot apply to the same event, a surpassing in (41), without contradiction:

(41) a. The man surpassed the boy of his childhood in regrets and not much else.
b. Many a man surpasses the boy of his childhood in regrets and not much else.

Only with supermonadicity providing enough events for the derived adverbs to modify will adverbialization license the substitutions. To substitute *the such-and-such* for *the so-and-so* in a sentence of natural language is always a double substitution in logical form—*the such-and-such while such-and-such-ing* for *the so-and-so while so-and-so-ing*. To be valid, it is not enough that the such-and-such are the so-and-so. It must also be that while such-and-such-ing is while so-and-so-ing, that is, that the derived adverbs describe the same events too.

§3. *Frames of reference and scenes.* The event that (2) describes and the events supermonadicity and adverbialization have so far introduced purport to deconstruct the goings on of the speaker's conscious report even as she is unaware of the fine-grained events enveloping her. Further annotations to logical form indicate perspective and frame of reference, representing the epistemic conditions for observation of the events reported. If, for example, the frame of reference for (42) is taken to be that of the ballerina's proximate visual experience, the sentence puts her partner in synchronous orbit around her (possibly leading from an outstretched arm):

(42) The ice ballerina pirouetted with her partner on the left across the rink.

Alternatively, the frame of reference could be fixed at a moment by the line-of-sight pointing across the rink and tangent to the ballerina's curved trajectory, tracking the partner in a parallel course, on one side or the other depending on whether the speaker faces the ballerina and the frame of reference is centered on the speaker's point of view or on the ballerina's. The sentence is thus three-ways ambiguous, and it communicates no determinate thought at all unless the speaker and hearer agree on the frames of reference intended. A lexical ambiguity—$left_1$, $left_2$ and $left_3$—to sort them out would insult the concept's integrity, which must rather be *left in f*, for different choices for frame of reference *f*. In contexts that are not demonstrative for *f*, as in third-party transmission,

to have understood (42) is to have grasped some descriptive intention as in (43), absent which, again, no determinate thought is expressed:

(43) In the orbital frames of references *f* centered on the agent's proximate visual experience, the ice ballerina pirouetted with her partner on the left in *f* across the rink.

Sentence (44) is also three-ways ambiguous—mercifully, not the nine ways it would be if the frame of reference for the second token of *left* were independent and possibly discontinuous from the first's:

(44) The ballerina will pirouette with her partner on the left to the middle of the ice, do a triple somersault, and pirouette with her partner on the left the rest of the way.

The instruction and knowledge not to switch frames of reference within sentence is itself a point of grammar. There is no thought here without description of a frame of reference, and its representation is subject to grammatical constraint.

For purposes of spatial orientation and navigation, analysis of the visual and auditory scene includes parameters justified by the agent's accurate navigation even as she is unaware of them in her experience of coarse-grained objects in location and in motion. Among the communicative intentions the design of language supports, I include the intention to convey to a companion witness accurate enough to guide her navigation, for which she has to acquire a scene from words of mine encoding parameters that I do not consciously attend to or assert, such as a frame of reference. How the world is at any given moment, as fixed in true propositions, is not accurate witness for where you are in it, for which the logical form of utterances should be tricked out to choreograph and calibrate the speaker's and hearer's thoughts and perceptions. (Think of two triangulating agents formulating and exchanging course corrections based on outputs and inputs in the language of a visual guidance system.) If narrative is rich enough to support the communicative intentions of its narrator, who is real or imagined (and always imagined if not real), it is shortsighted to suppose that the only parameter for narrative continuity is temporal—a timeline for world history and a succession of points or intervals along it as the narrative advances its 'now'. Even a 'here and now', the indices for a context of utterance, will not fix the thoughts expressed in a spatial vocabulary including *left*. Here we now are, imagine, at the moon landing site on schedule, and yet we are disoriented and disabled for visual navigation if the faulty periscope from our windowless rover swivels about randomly, no matter how detailed the image transmitted nor acute my report of it. A sequence of scenes or their narration is orienting and navigational for observer, speaker, or hearer only if their lines-of-sight are known and the scenes constitute a survey conforming to some natural conditions on their continuity—conditions that regulate over the course of survey and narrative such parameters as needed to constitute a natural (stereoscopic, i.e., 3-D) cinematic experience. As with any contextual or discourse parameters, thoughts are more or less dependent on them, and some narratives may wash out that dependency in adopting an ominiscient narrator or an arbitrary, unknown or neutral perspective. Nevertheless those parameters are always there for the asking.

In resolving a celebrated instance of the substitution puzzle, the metaphysics is spared only when conditions step in on the speaker's epistemic stance towards the events under report, granting that the nominal description contains expressions of such conditions:

(45) Three million passengers crowded National Airlines' routes last year.
(46) ⊭ Three million persons crowded National Airlines' routes last year.
(after Gupta 1980: 23; Moore 1994)

Frequent flyers recounted under (45)'s protocol keep it from implying (46). If the arithmetic predicate *three million* is a simple property '3M(_)' determined by the numeric identity of what is referred to, *three million passengers* does not refer to the one million whole persons who flew frequently. Away we go, denying the identity of passengers and persons to thwart the inference from (45) to (46) and asserting it in (47):

(47) The three million passengers who crowded National Airlines' routes last year were the one million frequent flyers loyal to it.

Worse, the three million passengers in (45) who are fewer persons do not always exist as such, subject to a condition that Doetjes & Honcoop (1997) aptly call *sequencing of events*:

(48) Three million passengers had three million opinions about the food on National Airlines.
(49) *Three million passengers have three million opinions about the food on National Airlines.

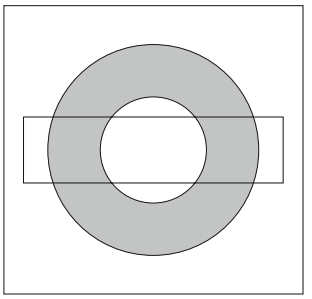
Sentence (45) can be followed by (48), recounting the passengers' experience onboard and continuing to count as many the fewer. In the present tense, however, (49) cannot count more passengers than persons who fly, despite the fact that these persons still have three million opinions about flights they remember all too well. The relation between the count and present tense wants explanation. Instead of the bare '3M(_)', imagine that *three million* translates as something like "now counted this way to three million," a description of events of measurement (the counter clicks, as it were) under an explicit protocol. What is counted and what the indefinite description *three million passengers* refers to is naively what there is, persons, who are counted and under some protocols recounted. With the counting now entered into the description, adverbialization derives (50) and(51):

(50) Three million passengers *while counted to three million* had three million opinions about the food on National Airlines.
(51) *Three million passengers *while counted to three million* have three million opinions about the food on National Airlines.

Clicking the counter as passengers go by frames or locates opinion recorded contemporaneously as in (48); but, there is no counting to three million that frames current conditions, which, according to (51), is what defeats (49). The speaker is secure in her constant reference to naive, familiar objects by a sleight of hand that manipulates her epistemic conditions, in this case, conditions of measurement.

A fundamental relationship exists between the conditions of measurement and the scenes of what is counted, which under adverbialization frame the events reported. In the narrow scene or frame of reference in (52), there are exactly two grey regions if finitely many. There is only one in the wide scene or frame of reference. This is not a metaphysical claim, for surely the regions counted under the narrow frame of reference still exist as parts of the grey ring under the wider frame of reference. They are just not to be counted there, under conditions inappropriate for their measurement.

(52)



Think of the design of an optical counter, blind to what is counted—passengers or planets if you say so. To utter *two* is to take on a perspective from which what is reported projects a scene the optical counter measures as 2. Counting two commits to the narrow scene or frame of reference in (52) and to a report of what the events counted two may frame within it, whether counting passengers or planets, as reflected in number agreement in the following:

(53) The morning star and the evening are playing hide-and-seek in the twilight, season after season.
(54) *From the vantage point of Jupiter, (you can see that) the morning star and the evening star are circling the sun.
(55) From the vantage point of Jupiter, (you can see that) the morning star and the evening star *is* circling the sun.

Narrow scenes while the morning star and while the evening star exist only from Earth. From Jupiter, if anything, what is while the morning star and while the evening star converges on the same large scene of Venus in orbit, where there is exactly one Venusian region, undermining the plural number agreement.

[Scenes are introduced with different purpose in (Schein 2002) to address puzzles of event identity, and perspectives in (Schein 1993: chapter 10) to complete the incomplete definite descriptions translating pronouns.]

§4. *Identity Statements* If supermonadicity and adverbialization are everywhere, they must also touch the most elementary assertions of identity. In any case, parsing (56)–(58) without them as closures of ⌜X = Y⌝ is a garden path into a substitution puzzle without exit, as the sentences could only chase after different values for X and Y:

(56) In a *mutual* declaration of *requited* love, the lover and belovèd in the first love note are *not* the lover and belovèd in the note returned.
(57) The lover and belovèd in either love note *are* the lovers exchanging them.
(58) ⊬ The lover and belovèd in the first love note are *not* the lovers exchanging them.

The logical forms and lexical items contained in (56)–(58) must admit modification by Tense at the very least and by the adverbs that, by hypothesis, adverbialization introduces with every nominal phrase. Beside hypothesis, modification in identity statements finds overt expression in (61), which differs from (60) in truth and differs only in the presence of a secondary predicate:

(59) Cezanne, Still Life with Vessels, Fruit and Tablecloth and Still Life with Ginger Jar, Sugar Bowl and Oranges

Photo credit: Scala/Art Resource, NY

Digital Image © The Museum of Modern Art/Licensed by SCALA/ Art Resource, NY

(60) T The oranges in the first still life are the oranges in the second.
(61) F The oranges in the first still life are the oranges in the second arranged the same way.

F They are them arranged the same way.

To be them is not to be them that way. It wants explanation how a classical identity statement *simpliciter* comes to be expressed in the same words as conditioned identity except for the condition itself. There is no room for modification in ⌜$X = Y$⌝. Instead, an identity relation '$\mathcal{I}[E_1, E_2]$' is said to hold between events or states only if the participants in E_1 are identical to the participants in E_2. $\mathcal{I}$dentity holds between the oranges' participation in the event or state depicted in the first canvas and their participation in that of the second canvas, verifying (60) (see (62)); but, that is not to say that their state in the second canvas is also the same arrangement. The further comment describing it as such results in the falsehood (61) (see (63)):

(62) [the X: ∃*E* oranges in 1st[*E*,X]][∃E_1: ∃*E* (*while*[*E*, E_1] & oranges in 1st [*E*,X])] participate[E_1, X] & [ɲE_1: *that*$_1$][ɲE_2: *that*$_2$] $\mathcal{I}$[E_1, E_2] &
[the X: ∃*E* oranges in 1st[*E*,X]][∃E_2: ∃*E* (*while*[*E*, E_2] & oranges in 2nd [*E*,X])] participate[E_2, X]

(63) [the X: ∃*E* oranges in 1st[*E*,X]][∃E_1: ∃*E* (*while*[*E*, E_1] & oranges in 1st [*E*,X])] participate[E_1, X] & [ɲE_1: *that*$_1$][ɲE_2: *that*$_2$] $\mathcal{I}$[E_1, E_2] &
[the X: ∃*E* oranges in 1st[*E*,X]][∃E_2: ∃*E* (*while*[*E*, E_2] & oranges in 2nd [*E*,X])] participate[E_2, X] & *arranged the same way*[E_2]

Given the common syntax that (60) and (61) warrant, adverbialization just joins in, as in (62) and (63), to further condition the events related by $\mathcal{I}$dentity. As weak as $\mathcal{I}$dentity is, (56), if it is just a negated identity *simpliciter*, falsely denies that there are events with the same participants in one of which the lovers participate while lover and belovèd and another in which they participate roles reversed. The sentence as intended is however to be understood as a true denial of a conditioned identity, where the condition is to be conveyed by an appropriately crafted, tacit secondary predicate:

(64) The lover and belovèd in a love note are *not* the lover and belovèd in the note returned *oriented the same way*.

(65) The lover *at the 1st address in a frame of reference*$_1$ and belovèd *at the 2nd address there*$_1$ are *not*
the lover *at the 1st address in a frame of reference*$_2$ and belovèd *at the 2nd address there*$_1$, *that frame of reference*$_1$ *oriented the same way as that frame of reference*$_2$.

If all the nominal phrases are addressed to a frame of reference, the secondary predicate is definable as a relation between the two frames of reference described, as the paraphrase in (65) hints at. Resolving this last class of substitution puzzles takes it all onboard—adverbialization, frames of reference, and the decomposition of identity statements to comport with supermonadicity. Yet, it respects that identity statements are always tensed and sometimes conditioned, while affirming that nominal phrases refer to what they naively refer to.

With all the event arguments in a supermonadic clause that adverbials, tense, and secondary predicates might modify, it becomes an empirical question to discover its grammar, which will lead to a canonical structure for the expression of spatiotemporal relations in a clause.

(66) For some evenings in 1892, Venus was Hesperus (aligned with a crescent moon).
For some evenings in 1892, Venus was the evening star (aligned with a crescent moon).

(67) * For some evenings in 1892, Hesperus was Venus (aligned with a crescent moon).
* For some evenings in 1892, the evening star was Venus (aligned with a crescent moon).

From the contrast (66)–(67), for example, it appears that the temporal frame adverbial, Tense, the secondary predicate, and the adverbial derived from the predicate nominal— in (66), *while Hesperus* or *while the evening star*— coincide in the events they describe, which are to be related by *I*dentity to events described by the subject's derived adverbial, *while Venus* in (66). Both modification and the asymmetry it induces in (66)–(67) again put the clause structure beyond reach of an analysis based on ⌜X = Y⌝.

§5. There is no rest for the Davidsonian who pauses at (2), as a few facts of natural language lead quickly to plural event quantification, plural event pronouns in lieu of bare variables, supermonadicity, and adverbialization. Along the way, the plural event pronouns are the means for *and* to be the univocal sentential connective in all its occurences, including (7) and (9), and all nominal phrases survive any context referring innocently to the familiar and ordinary, despite the puzzles of substitutivity. Metaphysical arguments to the contrary all rest on errors of syntax, parsing the puzzles without benefit of the enhanced clausal architecture. Some of these arguments are also deceived by the smoke and mirrors of scenes and frames of reference by which thoughts are calibrated and which logical form records with parameters for yet more events.

[I could not have attempted Event semantics while standing on one foot without leaning on Cleo Condoravdi, Richard Larson, Paul Pietroski, Roger Schwarzschild and Achille Varzi. I am grateful to Terje Lohndal for extensive correspondence on an earlier draft, and special thanks to Alexander Williams for written comments and subsequent correspondence about the final draft and to Elena Herburger for corrections minutes before midnight. The article draws on my (2010) '*And*': *Conjunction Reduction Redux*, USC, ms. 750pp.]

References

Alex Barber. 2000. A pragmatic treatment of simple sentences. *Analysis* 60.4: 300–308.

Greg N. Carlson. 1982. Generic terms and generic sentences. *Journal of Philosophical Logic* 11.2: 145–181.

Hector-Neri Castañeda. 1967. Comments. In Nicholas Rescher, ed. *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.

Donald Davidson. 1967. The logical form of action sentences. In Nicholas Rescher, ed. *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.

Donald Davidson. 1985. Adverbs of action. In Bruce Vermaze & Merrill Hintikka, eds., *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press.

Jenny Doetjes & Martin Honcoop. 1997. The semantics of event-related readings: a case for pair-quantification. In Anna Szabolcsi, ed., *Ways of Scope Taking*. Dordrecht: Kluwer Academic, 263–310.

Mürvet Enç. 1986. Towards a referential analysis of temporal expressions. *Linguistics & Philosophy* 9.4: 405–426.

Graeme Forbes. 1997. How much substitutivity? *Analysis* 57.2: 109–113.

Graeme Forbes. 1999. Enlightened semantics for simple sentences. *Analysis* 59.2: 86–91.

Jean Mark Gawron & Andrew Kehler. 2002. The semantics of the adjective *respective*. In L. Mikkelsen and C. Potts *WCCFL 21 Proceedings*, eds. Somerville, MA: Cascadilla Press, pp. 85–98.
Jean Mark Gawron & Andrew Kehler. 2004. The semantics of respective readings, conjunction, and filler-gap dependencies. *Linguistics and Philosophy* 27**:** 169–207.
Brendan Gillon 1990. Plural noun phrases and their readings: a reply to Lasersohn. *Linguistics & Philosophy* 13: 477–485.
Anil Gupta. 1980. *The Logic of Common Nouns: An Investigation in Quantified Modal Logic*. New Haven: Yale University Press.
Angelika Kratzer. 1996. Severing the external argument from the verb. In Johan Rooryck & Laurie Zaring, eds., *Phrase Structure and the Lexicon*. Dordrecht: Kluwer, pp. 109–137.
Angelika Kratzer. 2003. *The Event Argument*. Forthcoming from MIT Press.
Ernest Lepore & Brian McLaughlin, eds. 1985. *Actions and Events*. Oxford: Blackwell.
Lawrence B. Lombard. 1985. How not to flip the prowler: transitive verbs of action and the identity of actions. In Lepore and McLaughlin (1985), pp. 268–281.
Steven Moore. 1994. Events, objects and eventual objects. In Erin Duncan, Donka Farkas & Philip Spaelti, eds., *The Proceedings of the Twelfth West Coast Conference on Formal Linguistics: Conference Held at UCSC 1993.*, Stanford: Center for the Study of Language and Information, pp. 577–590.
Terence Parsons. 1985. Underlying events in the logical analysis of English. In Lepore and McLaughlin (1985), pp. 235–267.
Terence Parsons. 1990. *Events in the Semantics of English*. Cambridge MA: MIT Press.
Paul M. Pietroski. 1998. Actions, adjuncts and agency. *Mind* 107: 73–111.
Paul M. Pietroski. 2000. *Causing Actions*. Oxford: Oxford University Press.
Paul M. Pietroski. 2005. *Events and Semantic Architecture*. Oxford: Oxford University Press.
Stefano Predelli. 1999. Saul, Salmon, and Superman. *Analysis* 59.2: 113–116.
Jennifer M. Saul. 1997a. Substitution and simple sentences. *Analysis* 57.2: 102–108.
Jennifer M. Saul. 1997b. Reply to Forbes. *Analysis* 57.2: 114–118.
Jennifer M. Saul. 1999. Substitution, simple sentences, and sex scandals. *Analysis* 59.2: 106–112.
Jennifer M. Saul. 2007. *Simple Sentences, Substitution, and Intuitions*. Oxford: Oxford University Press.
Barry Schein. 1993. *Plurals and Events*. Cambridge, MA: MIT Press.
Barry Schein. 2002. Events and the semantic content of thematic relations. In Gerhard Preyer & Georg Peter, eds., *Logical Form and Language*. Oxford: Oxford University Press, pp. 263–344.
Barry Schein. 2006. Plurals. In Ernest Lepore & Barry C. Smith, eds. *The Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press, pp. 716–767.
Roger Schwarzschild. 1991. On the Meaning of Definite Plural Noun Phrases, unpublished Ph.D. dissertation, University of Massachusetts, Amherst.
Roger Schwarzschild. 1996. *Pluralities*. Dordrecht, The Netherlands: Kluwer.
Zoltán Gendler Szabó. 2003. On qualification. *Philosophical Perspectives 17(1)*: 385–414.

2.9

SKEPTICISM ABOUT MEANING

Michael McDermott

The several varieties of "meaning scepticism" reject a variety of plausible claims about meaning (not just about *knowledge* of meaning). The most important meaning skeptic is Quine. I shall also consider arguments of "Kripke's Wittgenstein" which have aroused some interest.

1 Quine

Quine held, roughly speaking, that all talk of meaning is meaningless. That makes it sound somewhat paradoxical, however. To understand Quine, we need some background.

First, Quine was an empiricist. 'Empiricism' is the name for a diverse collection of views united more by family resemblance than by deduction from any central tenets. It is useful, however, to distinguish two major strands.

(i) The Empiricist Theory of Justification is concerned with the foundations of knowledge, or of rational belief. It says, roughly speaking, that the foundation is observation, experience.

What other foundation could anyone suggest? Rejected candidates include divine revelation, Aristotle, the Church. The classical debate between empiricism and rationalism concerned the claims of "reason" to be a source of knowledge or, more properly, an *independent* source of knowledge (since it could hardly be denied that reason can provide knowledge when combined with experience—two persons who have had exactly the same experience can differ in knowledge). But the current eclipse of empiricism has been mainly at the hands of "common sense," particularly our alleged common-sense knowledge of *the mind* and *modality*.

"Observation," for the most significant empiricists (including Quine), was understood to be always of *publicly* observable events. There is, on this view, no non-inferential knowledge of the mind, no "introspection": if knowledge of the mind is possible at all, it is by inference from observations of behavior. Admittedly, some moderate empiricists were more liberal. And in much empiricist writing there is also some vestigial feeling that "observation" of the external world is itself really inference from internal data of some kind. This tension within empiricism was never satisfactorily resolved. But certainly there is a well established association between empiricism and "behaviourism".

The rejection of modal data was, by contrast, completely straightforward, for empiricists: observation is always of *actual* events.

(ii) The Empiricist Theory of Content says, roughly speaking, that the content of a statement is to be found in the same kind of thing that would make it reasonable to believe—experience. It's not quite that content = evidence, as it is sometimes put: Evidence of course consists of propositions about *past* experience, whereas content is a matter mainly of *future* experience; but they are the same *kind* of thing.

This is not the same as the ETJ, and has generally been found less plausible. Many "empiricists" are content with an uncritically commonsensical conception of content. They argue from empirical evidence for claims which seem to lack content on the ETC—for example claims about gods, or an unobserved past, or laws or minds conceived of in thoroughly "realist" terms. But the most significant and most controversial feature of the empiricism of Quine, and of his Vienna Circle predecessors, was its acceptance of a version of the ETC. This was *verificationism*: in its "classical" form it defined Viennese "logical positivism"; in its "holistic" form it was the center of Quine's philosophy.

Verificationism says that the content of a statement (positivism) or theory (Quine) is determined by what experiences would confirm or disconfirm it. If two statements/theories would be confirmed and disconfirmed by the same experiences, they have the same content.

The word 'content' is not exactly self-explanatory. It is often used, in empiricist writings, more or less interchangeably with 'meaning'. But it would be natural, when precision matters, to agree that they are not quite the same thing. It is natural to think that the content of a statement/theory is *determined by* the meanings of the words (and constructions) it uses. But two statements can have the same content even if they contain different words, with different meanings, e.g. 'John is taller than Mary' and 'Mary is shorter than John'. They are equivalent *by* meaning, though not identical *in* meaning; they are "synonymous" in a *broad* (and somewhat artificial) sense only. Similarly it is natural to say that "analytic" sentences, like 'Pigs are pigs', lack content, although of course they do not lack meaning.

Now how does Quine's empiricism relate to his views about meaning? The *first* key to understanding Quine, I suggest, is to see him as making an even bigger divorce between content and meaning than that just described. In these terms, Quine's view is that content is not even *determined by* meaning: a theory (not a statement) can have content without this content being determined by the *meanings* of its theoretical terms. The content of a theory is, for Quine, its predictions of experience. His point is that, in deriving a theory's implications for experience, recourse to knowledge of meanings will be unnecessary because the required "truths of meaning" will be explicitly stated (in a properly regimented theory). For example, deriving '*x* is not green' from '*x* is red' will require no knowledge of the meaning of 'red' or 'green' because 'Nothing red is green' will be available as an extra premise. The theorist can derive '*x* is not green' from '*x* is red' and 'Nothing red is green' without caring whether 'Nothing red is green' is a "truth of meaning".

This, I think, is the most important thesis of "Two Dogmas." It is, however, a thesis of the *epistemological irrelevance* of meaning, rather than the *nonexistence* of meaning. Theory construction is thought of by Quine as being done by an *individual*. Even if meaning is not relevant in that process, you might think it would be fundamental to the use of language for *communication*. On our main topic, the nonexistence of meaning,

"Two Dogmas" has much to say; but the argument is rather weak (in my opinion), in part because the *social* aspect of language is completely ignored.

The *second* key to understanding Quine is to see that he does have a good argument for the *nonexistence* of meaning—not in "Two Dogmas," but in *Word and Object*. And the argument rests on the ETC. The central question becomes, 'What is the *empirical content* of talk about meaning?' Quine assumes that it will have something to do with the *behavior* of language users—a *community* of language users. An adequate empirical theory for this domain, he argues, will not require anything like the common-sense meaning concepts. Furthermore, the common-sense concepts could contribute no additional empirical content. Words like 'meaning', 'analytic', and 'synonymous' will have no place in a properly scientific theory of language. In the end, what's wrong with talk of meaning is that it has no *content*.

So much by way of introduction. I now consider the two main Quine texts in more detail.

1.1 "Two Dogmas"

To lead up to Quine's arguments against meaning in TD, let me say something specifically about analyticity. Quine thinks of 'analytic' as being representative of the whole family of common-sense meaning concepts: It is straightforwardly interdefinable with 'synonymous' etc. But it does have a somewhat technical sound, so some preliminary explanation is in order.

For nonphilosophers, the purpose of language is to express our thoughts. Meaning is, on this view, a normative relation between sentences and thoughts: The meanings of the words in a statement determine whether or not you should assent to it, *given your beliefs*. Sentence content is inherited from belief content. Beliefs stand in relations of containment: The belief that $p\&q$, for example, contains the belief that p—there is no such state of mind as believing $p\&q$ but not believing p. Derivatively, there are relations of meaning containment between sentences: For example the content of the sentence $p\&q$ includes that of p—if you assent to $p\&q$ while dissenting from p, this cannot correctly express any genuine state of mind but must involve a misuse of language. Similarly, there seem to be sentences which cannot be assented to without misuse of language, for example '$p\&q$ but not p'. And others which cannot be *dis*sented from without misuse of language, for example 'not-($p\&q$ but not p)'. Assent to such sentences is required merely by the meanings of words, not by constraints of rationality or whatever pertaining to the underlying nonlinguistic states of mind. These are the "analytic" sentences. Let us say, to preserve connections, that $p\&q$ "analytically entails" p.

(Common sense is undecided, however, as to just how far these analyticities run, because it is undecided about the relevant psychology. You can't believe $p\&q$ without believing p. But can you believe that something is red without believing that it is not green? Can you believe that all men are mortal and Socrates is a man without believing that Socrates is mortal? Can you believe that everybody loves a lover and John loves Mary without believing that everybody loves *Mary*? (NB lover ≠ lovee.) If we interpret empiricist talk of analyticity in this natural psychological way, its claim against rationalism would be that *all* these entailments are analytic; *all* apparent use of deductive reason is merely verbal; deductive reason provides no new knowledge, it is just the discovery of new ways of expressing old beliefs; all a priori "knowledge" is analytic. But this seems to go way

beyond common sense. The rationalism versus empiricism issue is, however, irrelevant for our purposes: both sides agree that talk of meaning and analyticity makes sense.)

I have been trying to explain 'analytic' on the assumption that people have beliefs (nonlinguistic). For most of the twentieth century philosophers were pretty much agreed, however, that common-sense mentalism cannot be taken for granted for purposes of philosophical clarification. Mental content was held to be *more* problematic than the meaning or content of sentences or theories. The most popular method of philosophical clarification was through the analysis of language, rather than of thought. And the most popular approach to language involved "truth conditions." On this approach meaning is a relation between words and *the world* (rather than beliefs). Analyticity is vacuity of the nonlinguistic conditions of *truth* (rather than assent). The usual formula is: analytic = true in virtue of meanings alone.

There is a rather obvious difficulty: 'everything is self-identical', for example, is analytic if anything is, but can we deny that part of the reason it is true is that everything is self-identical? Would it be true if everything was *not* self-identical?

This truth-conditional definition of 'analytic' is Quine's initial target in TD. He does not criticize it directly, perhaps relying on the objection above. He considers a variety of attempts to improve on it. But we might wonder whether this is too quick: could 'true in virtue of meanings alone' actually be made to work? Let me digress to consider a couple of possible strategies.

(i) The ontologically committed approach. The meaning relation between words and things generates an association between sentences (*some* sentences) and non-linguistic "facts". Whether a sentence is true is normally a matter of whether there is a corresponding fact. But certain sentences are true without need of a corresponding fact: These are the analytic sentences. If we had a good theory of facts, and it said that there is no such fact as the fact that everything is self-identical, that would explain why 'Everything is self-identical' is analytic. Unfortunately, there appears to be no good theory of facts—not by empiricist standards, anyway. You could try saying that there is a fact that *p* iff *p* is *contingently* true, or knowable *a posteriori*, but this would make 'analytic' trivially equivalent to 'necessary', or to 'a priori', thus undermining the empiricist's claims against rationalism.

(ii) The ontologically uncommitted approach (Davidson). The idea here is that 'true (in L)' can do by itself as the central concept of a theory of meaning for L. The theory will have theorems like:

(1) 'Schnee ist weiss' is true (in G) iff snow is white.

Since our *total* theory will also contain

(2) Snow is white

and

(3) Grass is green,

and since its 'iff' will presumably be the *material* biconditional, it will also contain

(4) 'Schnee ist weiss' is true (in G) iff grass is green.

But it is plausible to suppose that within our total theory we may distinguish in some clear and principled way a subtheory, 'the semantics of G', which excludes (2), (3), and (4). Example (1) is thus a "truth of meaning"; (4) is not. (Quine, somewhat similarly, has no qualms about distinguishing a subtheory called 'logic'. Or we may distinguish 'biology' or 'genetics'. This is not philosophically contentious. There is no implication that the truths so singled out are true in some special way.)

Many truths of meaning are biconditionals. But perhaps some *unconditional* truths about truth in G, such as (5), will also be theorems of the semantics of G.

(5) 'Schnee ist weiss oder Schnee ist nicht weiss' is true (in G).

Would that make 'Schnee ist weiss oder Schnee ist nicht weiss' "true in virtue of meanings alone", i.e. analytic?

The trouble is that any subtheory presumably has as theorems *all* the logical truths expressible in it. So this approach too would trivialize the contentious claim that the truths of logic are all analytic.

Returning now to TD: Quine considers a variety of attempts to improve on 'true by meanings alone'.

The first of these is Frege's definition—S is analytic iff it is either a logical truth or can be turned into a logical truth by replacing words by synonymous words or phrases. Once again, the obvious objection is that the definition trivializes the claim that the truths of logic are all analytic.

Quine, however, has a different (but compatible) objection: the definition is *circular*. For 'synonymous' is "no less in need of clarification" than 'analytic'. Quine then considers certain attempts to define 'synonymous'. In particular he considers the suggestion that 'x is synonymous with y (in L)' means that x and y are intersubstitutable *salva veritate*. Quine observes that this suggestion will only work if L contains a specially tailored notion of "necessity" according to which 'necessarily p' is true iff p is analytic. The circularity is then obvious.

In general, the allegation of circularity is Quine's main weapon against the various definitions of 'analytic', 'synonymous' etc. considered in §§1–4. Since Quine does not tell us which terms he would *not* regard as "no less in need of clarification," this part of TD is not very impressive.

Section 5 introduces several new ideas: verificationism; a verificationist definition of 'analytic'; holism.

Verificationism, in its classical form, says that the way the meaning of the words in a statement determines its content is by determining that certain experiences would confirm it and certain other experiences would disconfirm it. An analytic statement is one whose meaning assigns zero content. (As opposed to "nonsense," where no definite content, zero or nonzero, is assigned.) Depending on how you understand 'confirm', this seems to imply that an analytic statement is one which (i) no experience would confirm, or (ii) every experience would confirm. If you understand confirmation as something like increase in probability, (i) would be the natural choice, since an analytic statement is presumably one whose meaning gives it maximum probability. A popular alternative was to take *dis*confirmation as basic, and defined in terms of analytic entailment: e disconfirms h iff h entails not-e; e would then confirm h if it did not disconfirm it. On this alternative (ii) would be more natural. (Here e is a *statement* reporting the occurrence of an experience.) Quine in TD seems to run the two together: an analytic statement

would be one which is "vacuously confirmed." He takes this as the verificationist definition of 'analytic'.

A natural objection to this definition would be that 'confirm', in the required sense, is no less in need of clarification than 'analytic'.

But Quine has a different objection: holism. In the ordinary sense of 'confirm', and presumably in any refined sense needed for verificationism, statements are not generally capable of confirmation; only theories are.

Holism is certainly a plausible and important thesis in its own right, but for our purposes the question is its relevance to the concept of meaning. Quine seems to think that it is crucial: since verificationism is basically correct, 'analytic' can only make sense if the verificationist definition can be accepted; but the definition fails on *holistic* verificationism.

I don't think this argument is any good. For holism does not prevent us from saying that the factual content of *theories* is determined by the meanings of their words. But if so much is granted, a notion of synonymy for words and sentences seems to be definable: *x* is synonymous with *y* iff replacement of *x* by *y* in any *theory* formulated in language L yields a theory with the same "empirical significance." And then S is analytic iff it is synonymous with 'Pigs are pigs'. Quine's target verificationist definition of 'analytic' is no good, certainly; but there is a *holistic* verificationist definition which the objection does not touch.

We should not leave TD without considering the paragraph of §6 which is most often quoted (including by Quine himself) as a summary of the key argument against analyticity. "Any sentence can be held true come what may," says Quine, and "no statement is immune to revision." Well, certainly we cannot define 'analytic' as 'immune to revision'—no defender of analyticity denies that you can deny an analytic sentence *if you change its meaning*. And certainly we cannot define 'analytic' as 'cannot be revised *without* change of meaning'—as a definition, that would be circular. But Quine seems to be suggesting something stronger: this description is not even *true* of any non-trivial class of "analytic" sentences—*any* sentence can be denied without change of meaning. This stronger suggestion, however, seems to be quite without support.

1.2 Word and Object

The treatment of linguistic meaning in Chapter 2 of WO is generally regarded as more radical than that in TD, and inferior. I disagree about "more radical." The conclusion in TD was utterly radical—all common-sense talk of meaning is nonsense. ('Analytic' was meant to be representative, not special.) As to "inferior," this perception has considerable justification. The central thesis of Chapter 2 is the "Indeterminacy of Translation," and I agree that this is a mess. It is not clear what the thesis is, nor what the arguments for it are supposed to be. (There is one reasonably clear and plausible argument, but only for a weaker thesis which came to be known as the Inscrutability of Reference.) Quine tried to do better later, but without success (on what it now a widely shared view). My advice is, forget the IT.

But there remains much in Chapter 2 that is not to be ignored. In particular, Quine's treatment of 'analytic' can easily be separated from the discussion of the IT. The difference from the treatment of 'analytic' in TD is striking. First, Quine's basic assumption in WO is behaviorism: he faces directly the question "Is there any empirical content in talk about meaning?"; and he assumes that, if there is, it consists of predictions of

observable behavior: namely the verbal behavior of a community of language users. Behaviorism didn't get a look-in in TD. Second, holism takes a back seat. In fact, Quine even seems to assume, contrary to holism, that the challenge for the defender of meaning is to reduce every individual statement about meaning to a prediction of behavior.

The argument against analyticity is that the best behavioral reduction of 'S is analytic' would be 'S is socially stimulus analytic', i.e. 'S would be assented to, under any conditions of current stimulation, by any speaker of the language'; but this is not close enough, since 'There have been black dogs' is socially stimulus analytic, but no defender of the ordinary notion would agree that it is analytic.

Quine realizes that a defender of the intuitive notion would protest that there is a better behavioral definition. An analytic sentence is, intuitively, one that we all assent to as a result of *shared linguistic habits*, not as a result of a shared *belief*. But Quine's behaviorism prohibits any reference to alleged mental *cause*s of verbal behavior: "All the objective data that [the linguist] has to go on are the forces that he sees impinging on the native's surfaces and the observable behavior, vocal and otherwise, of the native." (28) In general, the WO objection to intuitive talk of meaning is that it relies on an unreduced mentalism.

Unreduced? Could we hope to show that talk of belief etc. is, on the contrary, reducible to talk of behavior? Quine's reply is that the only plausible analysis would be along the lines of Carnap—'John believes that *s*' = 'John would assent to a sentence of his language *synonymous* with *s*'. It seems that we can only reduce meaning to behavior if we take mental states for granted; but we can only reduce mental states to behavior if we take meaning for granted; so neither meaning nor mind can be reduced to behavior. This is the "intentional circle." Given his behaviorism, Quine's conclusion is that there is no content in talk of meaning *or* in talk of mind.

This seems to me to be a more interesting argument against meaning than anything in TD. Of course it falls to the ground if you think the empiricist challenge to mind and meaning can be met by appeal to "common sense". Short of that, you could try to get out of the intentional circle by showing that talk of meaning has behavioral content without relying on anything mental; many have tried to do this. Or you could try to establish a behavioral foundation for mind talk and then use that to explain meaning; this strategy also has had occasional supporters. Or, finally, you could hold that there is *no* order of priority (Davidson). As things stand, it is not obvious which of these three strategies will ultimately be successful—if any. But certainly the prospects of success will be better if we take *holism* seriously: 'meaning', 'analytic', 'synonymous' etc., or perhaps 'belief', 'desire' etc., or perhaps all of these together, may be *theoretical* terms, in a theory with real empirical (i.e., behavioral) content—*reduction* is not necessary.

2 Kripke's Wittgenstein

KW's scepticism (Kripke 1982) focuses on the expression '. . . means . . . by . . .': he argues that it "makes no sense."

From a common-sense standpoint, this form of words is not unambiguous. Suppose you call my performance "meretricious"; I say "I presume you meant 'meritorious'." I might be suggesting that you meant (= intended) to *say* 'meritorious' (but your tongue slipped); this use of 'meant' is clearly irrelevant. Or I might be suggesting that you thought that 'meretricious' *means* 'meritorious'. This is not KW's target, either. His target is a supposed use of '. . . means . . . by . . .' which is *not* thus parasitic on the use

of 'means' in relation to a shared public language. His target is a concept of meaning applicable to a "private language" as well as to a public language.

KW's argument is that no *fact* can possibly do what meaning does (according to common sense). It proceeds in two main stages.

2.1 Stage 1

Common sense holds (says KW) that meaning something by a symbol is a state which *directs* the speaker how to use that symbol. For instance meaning addition by '+' is a state which directs the speaker to assent to '68+57=125'; meaning quaddition by '+' is a state which directs the speaker to assent to '68+57=5'. (Quaddition, standard symbol '⊕', is the function which differs from addition only in that its value is 5 for arguments greater than any relevant to the speaker's previous linguistic behavior.) What might constitute such a state?

The two leading candidates would appear to be

(i) A rule or intention to conform to certain remembered verbal instructions on how to use '+';
(ii) A rule or intention to "go on in the same way" as in certain remembered examples of the use of '+'.

But suggestion (i) is unsatisfactory because it just throws the problem back onto the words in the instructions: Unless there is a determinate fact that the speaker means such-and-such by *them*, following the instructions won't require any determinate verbal behavior. And suggestion (ii) is unsatisfactory because such a rule or intention is similarly indeterminate: Its content depends on the speaker's individual sense of analogy.

The argument so far seems to be very weak. Even without disputing anything substantial in KW's assumptions, it is easy to suggest a better candidate: meaning addition is a *complex* state of mind whose components are intention (ii) and the *usual* sense of analogy. Or, better, we can get rid of the implausible appeal to remembered instructions or examples: Someone means addition iff his rule or intention is to assent to '$n_1 + n_2 = n_3$' iff n_3 is the *sum* of n_1 and n_2. Or why not say, even more simply, that the difference between someone who means addition by '+' and someone who means quaddition is that only the former *means addition*? Why is that not an acceptable description of a genuine mental state as it stands? KW emphasizes that "there are no limitations . . . on the facts that may be cited to answer the sceptic." (14)

2.2 Stage 2

KW argues, now, that no fact can constitute meaning addition by '+' because meaning stands in a *normative* relation to verbal behavior. If I mean addition, then I speak *wrongly* if I assent to '68+57=5'. But it seems that no *fact* can justify, or make wrong, a bit of behavior. (No 'ought' from an 'is'.)

KW seems to accept that if we could find a *rule* or *intention* with a suitable content, it would have the required normativity. But this just shows that the real puzzle is how any fact could constitute a rule or intention.

I think this argument may be touching on something interesting and important. But it is not much of an argument for meaning skepticism as it stands. Not all uses of 'wrong' imply genuine normativity. Perhaps the clearest case is simple factual error. If John says the cat is on the mat and I say he's right and you say he's wrong, we just disagree whether the cat is on the mat. The question is, do rules and intentions have *genuine* normativity? I think we need at this point to distinguish between rules and intentions. Granted, if I intend to give you a dollar, and accidentally give you two, I give you the wrong amount. But this is not a good argument for the nonfactuality of intentions. It just shows that, in a suitable context, 'wrong' can mean 'unintended'. But perhaps rules, and in particular rules of meaning, do have genuine normativity. However, KW appears to be completely uninterested in the *difference* between rules and intentions.

You can *intentionally* disobey a rule. Perhaps this is the real reason that a "private language" (even one used for talking about publicly observable things) is impossible: The words of a personal shorthand cannot have *meaning*, in the ordinary sense, because there is no difference between using a word *contrary* to its meaning and *changing* its meaning. Perhaps, in general, there cannot be a "private rule" about anything. If I "make it a rule" to always brush my teeth after meals, and one day deliberately leave them unbrushed, have I *broken* the rule or *abandoned* it (perhaps replacing it with a less stringent rule)? If there is no difference, it would seem that it was not really a rule, in the ordinary sense, but merely an intention.

2.3 Skeptical Solution?

KW's meaning skepticism consists in the claim that there is no such *fact* as meaning addition—there is nothing which could make 'Jones means addition by '+' ' true, while preserving its normative implications. KW proposes, however, a "skeptical solution" to this "paradox." He proposes an alternative general picture of language. We should understand language by asking, not for truth conditions, but for *assertability* conditions. Under what conditions is it *justifiable* to assert *p*? What is the utility of having a sentence with such assertability conditions in our language? And he claims that, on this approach to language, the target locution *does* have meaning—as long as the speaker referred to is regarded as part of a community. 'Jones means addition by '+' ' can be justifiably asserted by Smith if Smith believes that Jones's use of '+' resembles his own. Similarly 'Jones means quaddition by '+' ' can be justifiably asserted by Smith if Smith believes that Jones's use of '+' resembles his own use of '⊕'. The utility of a sentence with such assertability conditions is that if Robinson hears Smith's remark, he will be able to infer that Jones's future use of '+' will resemble *his* (Robinson's) use of '+' or '⊕'; the possibility of this inference depends on community agreement—Robinson must be able to assume that *Smith*'s usage resembles his own.

It is hard to see why KW calls this account of '. . . means . . . by . . .' a "skeptical solution." For since the replacement of truth conditions by assertability conditions is offered as a *general* picture of language, applying equally to such paradigmatically factual sentences as 'The cat is on the mat', there seems to be nothing *skeptical* about it. And since it does nothing to show how 'Jones means addition by '+'" can have normative implications (when 'Jones's future use of '+' will resemble mine' does not), it does not seem to be a *solution* to the paradox.

References

Kripke, S. (1982) *Wittgenstein on Rules and Private Languages*, Cambridge: Harvard University Press.
Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.
Quine, W.V. (1961) "Two Dogmas of Empiricism" in *From a logical Point of View*, New York and Evanston: Harper and Row. (First published 1951.)

Section III

FOUNDATIONS OF SEMANTICS

3.1
NAMES

Barbara Abbott

John Stuart Mill regarded all definite noun phrases (NPs) as names, giving *the place which the wisdom or policy of antiquity had destined for the residence of the Abyssinian princes* as an example of a complex name (1843: 16). However, nowadays most people take the term "names" as equivalent to "proper names"—just NPs like *Madonna* and *Barack Obama*. Proper names have caused a lot of headaches for semanticians and philosophers of language, prompting David Kaplan to remark "if it weren't for the problem of how to get the kids to come in for dinner, I'd be inclined to just junk them" (1978: 224). This chapter will try to make clear what the main problem with proper names is, what solutions have been proposed, and why none of those solutions has been universally adopted as the correct one.

The Problem

The main problem with proper names concerns their contribution to the meaning of sentences (or utterances) in which they occur. (There are other problems, many of them related to this one, but we won't have space to go into them here.) Saying more about this main problem requires a bit of background. We'll start by looking at a different kind of definite NP—definite descriptions—which we will compare and contrast with proper names. Definite descriptions are NPs which (in English) begin with the determiner *the*. They are linked to their referents by expressing properties of those referents: *the star of Evita* refers to Madonna because she has the property of having starred in the movie *Evita*, while *the mother of Lourdes* refers to Madonna because she has the property of being the mother of Lourdes. Even though those two definite descriptions (*the star of Evita* and *the mother of Lourdes*) refer to the same person (Madonna), they do it by expressing different properties that Madonna has. These referent-determining properties are called "senses," after Gottlob Frege (1892).

The senses of definite descriptions obviously contribute to the meanings of sentences in which they occur—the propositions those sentences express. They do not, however, usually matter to the truth value of simple sentences. (1a) and (1b) below differ only in that one uses the NP *the star of Evita* and the other uses *the mother of Lourdes*.

(1) a. The star of *Evita* will be at the party.
    b. The mother of Lourdes will be at the party.

These two sentences obviously mean different things—they express different propositions. But it is clear that they have the same truth value; if one is true, then the other

must be also. Indeed this is exactly what we would expect, since how we refer to some entity shouldn't matter to the truth of what we say about it. This generalization is the law of substitutivity (also called "Leibniz's Law"), indicating that we would expect to be able to substitute one NP for a coreferential one without affecting the truth value of the containing sentence.

However, there are significant groups of sentences in which one can't exchange one definite description for another with the same referent and be sure of maintaining the same truth value. We focus here on one such group, which consists of sentences about mental attitudes such as knowledge, belief, desire, hope, anticipation, etc. So, it seems that (2a) might be true while (2b) is not.

(2) a. Frank knows that the star of *Evita* will be at the party.
b. Frank knows that the mother of Lourdes will be at the party.

The standard view is that belief, knowledge, desire, etc. are attitudes towards propositions—hence the term "propositional attitudes" for them. In the case of (2) the propositions in question are the ones expressed by the embedded sentences following *knows that* (these are the same as the sentences in (1)). And given that the properties expressed by a definite description contribute to the proposition expressed by the sentences it occurs in, it is not surprising that we can have a change of truth value when we exchange coreferential definite descriptions. The proposition which Frank is said to know in (2a) is different from the one he is being said to know in (2b).

We turn now to proper names. Unlike definite descriptions, proper names do not seem to achieve their reference by expressing properties of the things they denote. In other words they seem to be nondescriptional, lacking a Fregean sense. This is true even for names that originally expressed properties. Mill pointed out that, although the English town of Dartmouth originally got its name because it lay at the mouth of the Dart River, if an earthquake should cause the river to change its course the town would still be called by that name (Mill 1843: 20).

If proper names do not have any associated descriptional sense, then exchanging coreferential ones in a sentence should not change the proposition being expressed. The two sentences in (3) should express the same proposition.

(3) a. Mark Twain was a writer of humor.
b. Samuel Clemens was a writer of humor.

And we expect to be able to substitute coreferential proper names for each other in propositional attitude sentences without affecting the truth value of those sentences. However, it's not clear that this is the case; it seems that (4a) and (4b) could differ in truth value.

(4) a. Mary knows that Mark Twain was a writer of humor.
b. Mary knows that Samuel Clemens was a writer of humor.

So our problem is to explain this apparent difference.

The Solutions

The problem sketched above has been a worry to philosophers of language for a long time. First, below, we outline some early proposals to solve it. We then review a seminal

work of Saul Kripke's which changed the landscape, and by and large forced people to look in other directions for a solution. Following that, we take up some more recent proposals, separated into three groups. We'll pause after the first group to look at another of Kripke's contributions, which presented new challenges. We finish with the final two groups of approaches. We will not be able here to review all the analyses which have been put forward; important contributions from Mark Richard (1990), Graeme Forbes (1990), and others have had to be omitted. Nevertheless it is hoped that the reader will get some idea of the range of possibilities.

Early Solutions

Both Frege and Bertrand Russell concluded from data like those given above that (contrary to our intuitions and Mill's argument about Dartmouth) proper names do express properties of their referents, similar to those expressed by definite descriptions. Frege suggested that the name *Aristotle* might express a property like being the student of Plato who taught Alexander the Great (1892, 58 n.), and Russell suggested that the name *Bismarck* might be equivalent to *the first Chancellor of the German Empire* (1918: 171). Given this approach we might propose that *Mark Twain* means something like 'the most famous American humorist', while *Samuel Clemens* expresses a different sense (although it's a little difficult to say what that one would be).

One problem with such a theory is that it does not explain the difference in impact between examples such as those in (5).

(5) a. The student of Plato who taught Alexander the Great was a student of Plato.
b. Aristotle was a student of Plato.

The sentence in (5a) sounds analytic—that is, true just in virtue of its meaning. However, the sentence in (5b) does not sound analytic. But if the name *Aristotle* expressed the description suggested by Frege, then (5a) and (5b) should have exactly the same meaning, and if one sounds analytic then the other should too.

Others have solved this problem by proposing a looser kind of descriptiveness for proper names. This approach is sometimes called a "cluster" theory. On this kind of view, we associate a number of descriptions with a proper name, not all of which actually need to hold of the denoted entity. John Searle is one follower of the cluster theory. (Cf. also Wittgenstein 1953: §79.)

> Suppose we ask the users of the name 'Aristotle' to state what they regard as certain essential and established facts about him. Their answers would be a set of uniquely referring descriptive statements. Now what I am arguing is that the descriptive force of 'This is Aristotle' is to assert that a sufficient but so far unspecified number of these statements are true of this object.
>
> (Searle 1958: 171)

Ultimately Searle concluded that only one of the descriptions need hold true for a name to refer. That being the case, we no longer have a problem with the examples in (5). On Searle's theory, the logical form of (5b) is something like (6).

(6) The entity who had at least one of the following properties: $X_1, X_2, \ldots, X_n$, had property X_m, where $1 \leq m \leq n$.

And sentences of the form of (6) are not analytic.

Kripke (1972) on Proper Names

In a major series of lectures titled "Naming and necessity" (later published as a book), Kripke (1972) gave a number of arguments against both single description theories and cluster theories, and in favor of Mill's conclusion that proper names are nondescriptional. He pointed out that people can know and use names for individuals when they don't know any identifying properties of those individuals. I've heard of Anaximander, for example, and can use his name to ask questions about him, but at present I don't know anything more about him than that he was an ancient Greek philosopher. Contrary to what the quote from Searle suggests, I could not give any uniquely referring descriptive statements.

Kripke also pointed out that when we speak of alternative possibilities—different possible worlds—definite descriptions differ from proper names in potentially referring to different entities. In the actual world the NP *the star of Evita* denotes Madonna, but in different circumstances the movie maker could have picked a different person for the starring role, and in such a possible world *the star of Evita* would refer to that person and not Madonna. Thus (7) is true.

(7) Madonna might not have been the star of *Evita*.

However, proper names seem to pick the same individual out of any possible world. Because they don't express a property of that entity, they have no way to change their referent under different possible circumstances. Thus (8) does not seem to be true.

(8) Madonna might not have been Madonna.

(In thinking about alternative possible worlds, and how expressions of our language apply to them, it is important to be clear that we are keeping our language fixed. The purpose of these exercises is to become clearer on meaning in our language as we speak it. Thus, it is not relevant that Madonna might not have been named that in a different possible world, or that the word *star* might not have been used for leading movie actors.)

Kripke called proper names "rigid designators," meaning that (as we use them) they refer to the same entity in all possible worlds (1972: 48). He proposed that instead of being connected to their referents by expressing properties of those referents, proper names are connected by a historical or causal chain—a chain of uses of the name going back to the original baptism or dubbing.

These facts present a problem for both the single description theory and the cluster theory of proper names. No matter which single description we pick, as long as it is a description of the type suggested by Frege, Russell, or Searle, it will not hold of the same entity in all possible worlds. (There are what we might call "necessary" descriptions, like *the sum of 5 and 7*, which apply to the same entity (12, in this case) in all possible worlds. However contingent descriptions of the kind we've been considering—like

the star of Evita—are not of this type.) On the other hand as we've just seen, proper names do pick the same entity out of all possible worlds. The sentences in (7) and (8) made this contrast clear. Furthermore, like (7), (9) seems also to be true.

(9) Madonna might have had none of the properties commonly attributed to her.

This is a problem for the cluster theory. If a proper name meant something like 'the entity with at least one of following commonly attributed properties: $X_1, \ldots, X_n$,' then (9) should sound contradictory, but it doesn't.

Most people who write about this subject have found Kripke's arguments convincing, and have looked elsewhere for a solution to the problem presented by proper names in propositional attitude sentences. Most of the proposed solutions fall into one of three main categories: metalinguistic approaches, which involve the names themselves in one way or another; so-called "hidden indexical" approaches, which introduce extra material into propositional attitude sentences; and what I'll call "bite the bullet" approaches, which try to explain the problem away. We'll start with a look at the metalinguistic approaches, and then introduce some more data which is particularly difficult for such approaches to account for, which will cause us to look at the other two kinds of potential solution.

Metalinguistic Approaches

There are three subcategories of metalinguistic approach to consider. The first involves a different analysis of propositional attitude sentences, while the second and third are new analyses of proper names.

Quotation

Several philosophers (e.g. Quine 1956; Davidson 1969) have suggested that propositional attitude verbs like *know*, *believe*, *hope*, etc. actually express relations to sentences rather than propositions. On this quotational view our sentences in (4) above would be analyzed as in (10).

(10) a. Mary knows-to-be-true "Mark Twain was a writer of humor."
b. Mary knows-to-be-true "Samuel Clemens was a writer of humor."

According to this analysis (4a) (= (10a)) attributes to Mary knowledge of the truth of the sentence *Mark Twain was a writer of humor*, while (4b) (= (10b)) attributes to Mary knowledge of the truth of a different sentence—*Samuel Clemens was a writer of humor*. Since two different sentences are involved, it is not a problem that one of these could be true and the other false.

One potential problem with quotational views is that they seem to suggest that any holder of a propositional attitude must also know a language (although it has been argued that this is not actually a consequence of such views). However we commonly attribute knowledge, beliefs, and desires to languageless creatures such as cats, dogs, chimpanzees, and prelinguistic children. Indeed, animals of all sorts are capable of performing intricate series of tasks in highly organized ways to obtain just the things they

need to survive, and it is difficult to imagine that they do this without a thought or a desire in their head.

Causal description

We saw above that Kripke had proposed that names are linked with their referents via a historical or causal chain of use (rather than a descriptive sense). A second group of metalinguistic approaches incorporates this causal chain into a special kind of descriptive meaning for a proper name. (See, for example, Devitt 1981, Searle 1983, Lewis 1997.) One potential problem for this view is that most people do not know anything about the historical chain leading up to their use of a name. In most cases they will have completely forgotten when and where they first heard it used. Yet they use these names confidently, as though nothing were lacking in their knowledge of them. (See Abbott 2010: 117–18, and the works cited there, for further discussion.)

'The bearer of N'

The third, and currently most popular, kind of metalinguistic approach holds that a name *N* expresses a meaning something like 'the entity named *N*' or 'the bearer of *N*.' (See, e.g., Katz 1977, 2001; Bach 1987, 2002; Recanati 1993; Geurts 1997.) Kripke had responded to an earlier proponent of this kind of approach (Kneale 1962) with several arguments. For one thing, Kripke thought it was circular: "We ask, 'To whom does he refer by "Socrates"?' And then the answer is given, "'Well, he refers to the man to whom he refers'" (1972: 70). And for another, Kripke pointed out that such a theory would not make sense for other words—we would hardly want to have a dictionary with entries like "*horse*: n., those things called 'horses'" (1972: 69). A later paper of Kripke's presented another problem, perhaps even more serious. We turn now to that.

Kripke's Puzzle

In "A puzzle about belief" (Kripke 1979), Kripke describes the following situation. Pierre is a monolingual French boy who, based on picture books and postcards, forms a belief which he expresses by uttering (11).

(11) Londres est jolie.

Unfortunately he is kidnapped, taken abroad, and forced into the slave trade in the very worst part of London (I've filled in a few details here), where he learns English by the direct method (that is, with no instruction). Now he is inclined to assert (12)

(12) London is not pretty.

although none of his attitudes have changed about the city he still calls *Londres*. Kripke asks the simple question: Does Pierre, or does he not, believe that London is pretty?

Kripke gave another version of the puzzle which does not require two languages. Peter has heard of Paderewski, the famous Polish statesman. He has also heard of someone with the same name who was a celebrated pianist and composer, but believes this to be a different person. He also thinks that people in politics are never musically inclined.

So the question is, does Peter, or does he not, believe that Paderewski had musical talent?

One thing shown by these examples is that the problem with proper names in propositional attitude contexts is not, or not only, one of substitution. In the Pierre case we have corresponding sentences in two different languages, and in the Paderewski case there is just one sentence of English involved. But Kripke believes that the examples show something much deeper, a fundamental problem with our attributions of propositional attitudes.

> When we enter into the area exemplified by . . . Pierre, we enter into an area where our normal practices of interpretation and attribution of belief are subjected to the greatest possible strain, perhaps to the point of breakdown. So is the notion of the *content* of someone's assertion, the *proposition* it expresses.
>
> (Kripke 1979: 269, italics in original)

It seems that in Kripke's view, the simple questions posed about Pierre and Peter are unanswerable.

What is the consequence of these examples for our examples in (4), repeated here?

(13) a. Mary knows that Mark Twain was a writer of humor.
b. Mary knows that Samuel Clemens was a writer of humor.

Kripke holds that the same considerations apply—that we cannot simply chalk up the seeming truth of (13a) and falsity of (13b) to a failure of substitutivity. Instead he has called into question the very possibility of attributing any propositional attitude involving a proper name.

Kripke's puzzles seem to present a problem for metalinguistic theories of proper names and/or propositional attitude contexts. Recall that those theories involve the name itself—either directly (as in the quotational approaches), or because of the historical chain that leads to its use (as in the causal description approaches), or in a description of the form "the bearer of *N*." But in Peter's case we apparently have just a single name, so it is not clear how any of these approaches can explain the difficulty in attributing to him a belief about Paderewski. Which is not to say that there have not been responses—the references cited in the preceding subsection which come after 1979 have all had to address Kripke's puzzle, and the interested reader should investigate these resources to see what their authors have to say.

We turn now to the two remaining categories of response to the problem of proper names in propositional attitude sentences.

Hidden Indexical Approaches

Generally speaking, hidden indexical approaches such as those of, for example, Crimmins and Perry (1989), and Schiffer (1992), accept Kripke's arguments that proper names are nondescriptional, and hence do not add descriptional content to the propositions expressed by sentences containing them. In order to solve the problem of proper names in propositional attitude sentences, they regard such sentences as containing additional constituents which are not overtly expressed—hence the adjective "hidden." Furthermore, the exact content of these constituents may depend on the context in which the

sentence is uttered—hence the word "indexical." Let us try to get clear on what this kind of approach involves.

The basic idea is first to recognize that people have conceptions of the things they have beliefs and other attitudes about, and then to postulate that one or another of those conceptions may be implicitly included in a propositional attitude report. These conceptions are sometimes called "modes of presentation" (using a term which originated with Frege (1892) but in a slightly different way from the way Frege used it). When we say, for instance, that Mary knows that Mark Twain was a famous humorist, we implicitly invoke a mode by which Mark Twain is psychologically presented to Mary and which would typically include facts about him such as his having written amusing books and made amusing comments.

Furthermore these theorists believe that these additional constituents, the modes of presentation, are part of the semantics of the utterance—they play a role in determining its truth or falsehood. Thus the logical form of a propositional attitude attribution will involve an additional argument place, not represented in the surface form of the sentence. We might represent the logical form of our examples about Mary, on this kind of view, as in (14).

(14) a. Knows (Mary, that Mark Twain was a famous humorist, m_1)
b. Knows (Mary, that Samuel Clemens was a famous humorist, m_2)

In both (14a) and (14b) Mary is said to be in the knowledge relation to the same proposition, i.e. the proposition that Mark Twain/Samuel Clemens was a famous humorist. What differentiates (14a) from (14b) are the distinct modes of presentation m_1 and m_2, and in order for these sentences to be true, the knowledge Mary is said to have must be via the correct mode of presentation. Presumably in a context where (14b) strikes us as false, it will be the case that the unexpressed mode of presentation is one which does not include the stereotypical facts that we associate with Mark Twain, the famous humorist.

In arguing for the existence of such unexpressed constituents, Crimmins and Perry draw an analogy to sentences about the weather (1989: 699–700). One can't say *It's snowing* and sure of saying something true simply because it's snowing somewhere or other. Instead, when we say *It's snowing* there is implicit reference to a specific place where the snow is said to be occurring—usually, but not always, the place where the speaker is located. In order for the speaker to have said something true, it must be snowing at that location.

Turning to Kripke's puzzle, hidden indexical theorists will typically hold that a sentence like (15)

(15) Peter believes that Paderewski had musical talent.

will express different things in different contexts, because the mode of presentation which is implicitly being attributed to Peter will be different. In a context in which Paderewski's musical accomplishments are salient, the mode will be one which includes that kind of thing, and (15) will be true. However, in a context where, perhaps, great Eastern European political figures are under discussion, a different mode of presentation will be invoked—one involving Peter's beliefs about Paderewski as a statesman. In that kind of context (15) expresses a different proposition, one which is false.

Nathan Salmon (1995) has raised a number of objections to the hidden indexical type of approach. (See also Richard 1993.) One concerns the result just cited—that (15) will be true on some occasions of its use and false on others. In its surface form, (15) does not have any relevant indexicals, so to determine that it is on some occasions true and on others false when none of the relevant facts about the world have changed is highly counterintuitive. A more fundamental objection is that the modes of presentation invoked by the hidden indexical theorists should not be considered to be part of the meaning expressed by propositional attitude sentences but should instead be attributed to pragmatic considerations. They are so treated in the last kind of approach to be considered—what I've called "bite-the-bullet" approaches.

"Bite-the-Bullet" Approaches

The last kind of theory of proper names in propositional attitude contexts we'll consider has been promoted largely by Salmon (e.g. 1986, 1995), and Scott Soames (e.g. 2002). (Cf. also Braun 1998.) This kind of approach accepts Kripke's conclusions concerning the nondescriptionality of proper names, and does not posit any additional hidden indexical content in propositional attitude ascriptions. What, then, does this kind of approach have to say about our original sentences about Mary, repeated here once again?

(16) a. Mary knows that Mark Twain was a writer of humor.
b. Mary knows that Samuel Clemens was a writer of humor.

Salmon and Soames acknowledge that holders of propositional attitudes will conceive of the propositions involved in various ways—what we've been calling "modes of presentation." However, they hold that specific modes are not represented in any way in a typical propositional attitude ascription such as those in (16). Instead, it is simply the case that one holds a particular propositional attitude if there is at least one way in which one both conceives of the proposition and has the relevant attitude toward it. If Mary has a way of thinking of Mark Twain according to which she knows that he was a writer of humor, then (16a) and (16b) are both true.

In order to account for our intuition that (16b) is false, Salmon and Soames point out that, under the circumstances, the speaker of (16b) has presented its content in a misleading way. Generally speaking, when one attributes an attitude to another, one will choose a form of words which reflects the way the attitude holder grasps the proposition in question, and addressees expect speakers to do that. By using the name *Samuel Clemens*, the utterer of (16b) suggests that Mary is familiar with Mark Twain under this name as well, but this suggestion is not an entailment of the utterance. In this way the bite-the-bullet approaches attribute modes of presentation to pragmatic inference from an utterance, and not to its semantic content as in the hidden indexical views.

When it comes to Peter and Paderewski, the bite-the-bullet approach says that (15) (*Peter believes that Paderewski had musical talent*) is simply true. It is true because there is at least one way Peter has of conceptualizing Paderewski under which he recognizes the Pole as an accomplished musician. Under this approach (17) would also be deemed true.

(17) Peter believes that Paderewski did not have musical talent.

That is because Peter has another way of conceptualizing Paderewski under which the Pole seems definitely not to have been a musician. This is not to saddle Peter with irrationality, however, since two different guises are involved.

Concluding Remarks

It is hoped that the preceding survey has given the reader some idea of the range of approaches to solving the problem of proper names in propositional attitude contexts. The very existence of so many different analyses of proper names, and of propositional attitude sentences, indicates the unfinished state of research in this area. In the background lies Kripke's challenge concerning Peter and Paderewski: Is it possible to come up with a clearly satisfactory analysis of proper names and propositional attitude sentences that will give a definite response to his question? Or is he right, that such attributions lie in an area "where our normal practices of interpretation and attribution of belief are subjected to the greatest possible strain, perhaps to the point of breakdown"? Speaking for myself, I incline toward the bite-the-bullet type of theory, but not everybody would agree.

Related Topics

1.1 Extensions, Intensions, Character, and Beyond
1.13 Empty Names
2.3 Propositions
3.7 Descriptions
6.7 Propositional Attitudes.

References

Abbott, Barbara (2010) *Reference*, Oxford: Oxford University Press.
Bach, Kent (1987) *Thought and Reference*, Oxford: Oxford University Press.
—— (2002) "Giorgione was So-called Because of his Name," *Philosophical Perspectives* 16: 73–103.
Braun, David (1998) Understanding Belief Reports. *Philosophical Review* 107: 555–595.
Crimmins, Mark and John Perry (1989) "The Prince and the Phone Booth," *Journal of Philosophy* 86: 685–711.
Davidson, Donald (1969) "On Saying That," in Donald Davidson and Jaakko Hintikka (eds.) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158–174.
Devitt, Michael (1981) *Designation*, New York: Columbia University Press.
Forbes, Graeme (1990) "The Indispensability of *Sinn*," *Philosophical Review* 99: 535–563.
Frege, Gottlob (1892) "Über Sinn und Bedeutung," *Zeitschrift für Philosophie und Philosophische Kritik*, 25–50. Translated as "On Sense and Reference", in Peter Geach and Max Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 56–78.
Geurts, Bart (1997) "Good News about the Description Theory of Names," *Journal of Semantics* 14: 319–348.
Kaplan, David (1978) "Dthat," in Peter Cole (ed.) *Syntax and Semantics, vol. 9: Pragmatics*, New York: Academic Press, 221–243.
Katz, Jerrold J. (1977) "A Proper Theory of Names," *Philosophical Studies* 31: 1–80.
——(2001) "The End of Millianism: Multiple Bearers, Improper Names, and Compositional Meaning," *Journal of Philosophy* 98: 137–166.
Kneale, William (1962) "Modality De Dicto and De Re," In Ernest Nagel, Patrick Suppes and Alfred Tarski (eds.) *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, Stanford: Stanford University Press, 622–633.
Kripke, Saul (1972) "Naming and Necessity," in Donald Davidson and Gilbert Harman (eds.) *Semantics of*

Natural Language, Dordrecht: Reidel, 253–355 and 763–769. Page references are to the 1980 monograph reissue, Cambridge, MA: Harvard University Press.

——(1979) "A Puzzle about Belief," in Avishai Margalit (ed.) *Meaning and Use*, Dordrecht: D. Reidel, 139–183.

Lewis, David (1997) "Naming the Colours," *Australasian Journal of Philosophy* 75: 325–342.

Mill, John Stuart (1843) *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*, London: John W. Parker. Page references are to the 1891 edition, reissued 2002, Honolulu, HI: University Press of the Pacific.

Quine, W.V. (1956) "Quantifiers and Propositional Attitudes," *Journal of Philosophy* 53, 177–187.

Recanati, François (1993) *Direct Reference: From Language to Thought*, Oxford: Blackwell.

Richard, Mark (1990) *Propositional Attitudes*, Cambridge: Cambridge University Press.

—— (1993) "Attitudes in Context," *Linguistics and Philosophy* 16: 123–148.

Russell, Bertrand (1918) "Knowledge by Acquaintance and Knowledge by Description," in *Mysticism and Logic*, London: Longmans, Green. Page references are to the 2004 republication, Mineola, NY: Dover, 165–183.

Salmon, Nathan (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.

——(1995) "Being of Two Minds: Belief with Doubt" *Noûs* 29: 1–20.

Schiffer, Stephen (1992) "Belief Ascription," *Journal of Philosophy* 89: 499–521.

Searle, John R. (1958) "Proper Names," *Mind* 67: 166–173.

——(1983) "Proper Names and Intentionality," in *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press, 231–261.

Soames, Scott (2002) *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, Oxford: Oxford University Press.

Wittgenstein, Ludwig (1953) *Philosophical Investigations*, translated by G.E.M. Anscombe, Oxford: Basil Blackwell.

3.2
VERBS

Zoltán Gendler Szabó

The earliest known Greek grammar by Dionysius Thrax distinguishes eight parts of speech: noun, verb, participle, interjection, pronoun, preposition, adverb, and conjunction. The list looks largely familiar to the modern eye, though there are certain surprises. A number of categories that play a role in modern taxonomies—article, auxiliary, classifier, complementizer, determiner, etc.—are missing. Some of the traditional categories appear too narrow: participles, prepositions, and pronouns are typically viewed by contemporary linguists as subcategories of adjectives, adpositions, and nouns, respectively. When it comes to interjections, there is a serious question whether they should count as genuine words at all: greetings, expletives, fillers, and the like are at best loosely integrated into sentences. But in the larger scheme, these are quibbles: we continue to classify words roughly the same way as we did around 100 B.C. Given that, it is striking that there is no agreement about the underlying principles of classification.

Among the usual categories, it's verbs that probably have the best chance of being instantiated in all human languages (Baker 2003: sections 2.20, 3.9, 4.6). Dionysius identifies verbs by a morphological and a semantic criterion: They are words inflected for person, number, and tense that signify actions or passions. This definition has been reprinted thousands of times with slight variations. It works remarkably well in practice—as generations of schoolchildren can attest. But if we are interested in what verbs are—as opposed to how to spot most of them in Ancient Greek—the traditional criteria won't do. Vietnamese, Thai, Mandarin Chinese, and a number of other languages have no system of verbal inflections. And no matter how far one stretches the meanings of "action" and "passion," there are plenty of counterexamples to the claim that all verbs signify one or the other: The Latin verbs *existit* [exist], *quiescit* [rest], *friget* [show cold], *albet* [show white], or *claret* [show bright] are but a few mentioned already in the *Port-Royal Logic* (Arnauld and Nicole 1662: 80). English lacks a morpheme like *-et* but such morphemes do exist in other languages besides Latin; cf. the German *grün+en* and the Hungarian *zöld+ell* [show green].

The problems with the traditional definition can perhaps be fixed. We can move from the purely morphological criterion to a more syntactic one and say that verbs are the only predicates in the lexicon. We can also sharpen the semantic criterion and say that verbs are true of events, processes, and states but never of other sorts of entities. A defense of these characterizations is presented in sections 1 and 2 below. The proposed definition—verb is a lexical category whose members can be predicated of events, broadly construed—remains silent on the two features that have been widely regarded as constituting the very essence of the verb. Aristotle says that "a verb is what

additionally signifies time" (*De Interpretatione* 16b6) and that "every statement-making sentence must contain a verb" (*De Interpretatione* 17a9). Though neither of these claims is strictly true, they contain important insights; these will be examined in sections 3 and 4.

1 Verbs Are Predicates

Officially, expressions belong to the same syntactic category just in case they can be interchanged within other expressions *salva beneformatione* (without loss of well-formedness) and they belong to the same semantic category just in case they can be interchanged within other expressions *salva significatione* (without loss of meaningfulness); cf. Bar-Hillel 1953. But the official line is routinely ignored because it leads to excessively fine-grained categories. Consider the bound morpheme *-ly*. It can be added to *hour, day*, *month,* and *year* but not to *second*, *minute*, *decade,* and *century*. But few would be willing to make a grammatical distinction on such feeble grounds. (It seems that *-ly* can combine with nouns denoting medium-length intervals only.) Linguistics is committed to the unity of the category of nouns, not only in the face of minor differences in what morphemes they combine with but despite massive differences in syntactic behavior between count nouns and mass nouns, common nouns and proper nouns, relational and nonrelational nouns, etc. The consensus is that words do different kinds of things in language and we regard distribution as a mere clue for discovering what these job-descriptions really are.

The prime division among words is between *lexical* and *functional* categories. The latter were traditionally seen as contributing to the form, rather than the content of sentences. Paradigmatic examples of functional words are the determiner *every*, the conjunction *or*, and the complementizer *that*. Functional words are short and tend not to be loaned; functional categories are closed and tend to be small. In addition, lexical categories do and functional categories don't participate in derivational morphology. For example, nouns verbs, adjectives can be transformed into each other in a variety of ways but not to adpositions. And there are no morphemes that can attach to adpositions that would turn them into a noun, a verb, or an adjective. This arguably shows that adposition is a functional category despite the clear spatial meaning many of its members have. (This is a contested claim: both Chomsky 1970 and Jackendoff 1977 count adposition as a lexical category. The Appendix of Baker 2003 gives a detailed argument against their view.)

There are apparent exceptions to the claim that functional categories fail to participate in derivational morphology. For example, in English many prepositions combine with *-ness*: cf. *betweenness*, *aboveness* and some auxiliaries combine with *un-*: cf, *undo*, *unhave*. But *between* and *above* have adjectival variants and *have* and *do* have verbal ones. Thus it seems reasonable to postulate ambiguity in these cases: perhaps there are disitinct words pronounced as /bɪ'twi:n/, /ə'b^v/, /du/, /hæv/.

Verb is clearly a lexical category; the question is what distinguishes it from other lexical categories. My suggestion is simple: Verbs, unlike other lexical categories, are predicates—expressions used to say something *of* something. Their predicative nature is reflected in the fact that they are associated with a set of argument positions. Nouns and adjectives can have argument positions too: *mother* and *proud* both have one marked by the preposition *of*. What sets verbs apart for these nouns and adjectives is that they yield a clause when their arguments are saturated.

Being a predicate is widely accepted as a necessary condition for being a verb, but it is equally widely rejected as a sufficient condition. The near-consensus is that a great many other expressions—in particular, common nouns and intersective adjectives—are also predicates. Such expressions can certainly be used predicatively, but they clearly have other uses as well. Common nouns can be used to refer (cf. *dog* in *Dogs are animals*) and intersective adjectives can be used to modify (cf. *large* in *Large dogs bark*). In principle, we could see either the predicative or the nonpredicative use as primary. By contrast, verbs have *only* predicative uses. A verb can come to refer only if it is transformed into a noun (for example a gerund) and it can come to modify only if it is transformed into an adjective (for example a participle). Thus, verbs *must* be counted as predicates, but nouns and adjectives *needn't* be. We get the tidiest picture if we assume that nouns and adjectives get to be predicates only as a result of a syntactic transformation.

There is evidence that predicative occurrences of nouns and adjectives contain extra structure. The first thing that comes to mind in this connection is that simple declarative sentences involving nominal and adjectival predicates require (at least in most languages) a verbal copula. But this is not conclusive: in small clauses (for example the embedded clauses of *I regard the senator a crook* and *You consider the senator lucky*) there is no overt verb linking the subject with a nominal or adjectival predicate. A better reason to believe that nonverbal predicates are more complex than verbal ones comes from coordination:

(1) We consider the senator lucky and a crook.
(1') We consider the senator $[_{\text{PREDP}}\ [_{\text{PR}}{}^{\varnothing}][_{\text{AP}}$ lucky]] and $[_{\text{PREDP}}\ [_{\text{PR}}{}^{\varnothing}][_{\text{NP}}$ a crook]]

Since the adjectival and nominal predicates can be conjoined they must be of matching categories. This is guaranteed if we assume that when adjectival or nominal phrases are used predicatively they are embedded into a predicate phrase. The idea can be implemented by positing an empty predication head PR taking an AP or NP as complement. The subject of predication may start in the specifier position and be subsequently raised out of the PREDP (Bowers 1993, 2001). The fact that verbs do not coordinate with nonverbal predicates suggests that they do not need such an empty element to become predicates (Baker 2003: 78):

(2) *His advisors made the senator win and lucky.
(2') *His advisors made the senator $[_{\text{VP}}$ win]] and $[_{\text{PREDP}}\ [_{\text{PR}}{}^{\varnothing}][_{\text{AP}}$ lucky]]

(3) *His advisors made the senator win and a crook.
(3') *His advisors made the senator $[_{\text{VP}}$ win]] and $[_{\text{PREDP}}\ [_{\text{PR}}{}^{\varnothing}][_{\text{NP}}$ a crook]]

It has been suggested that PR is lexically realized in English by the *as* particle in *We regard the senator as lucky* and *We regard the senator as a crook* (Bowers 1993). Many other languages have particles that arguably realize PR: *3a* in Russian, *som* in Norwegian, *als* in German, *yn* in Welsh, *ina* in Irish, *-kye* in Korean, etc. Some of these particles take only APs or NPs. Edo apparently has distinct particles for constructing nominal and adverbial predicates: *yé* performs the former function, *rè* the latter (Baker 2003: 40). Overt PR is the norm for languages throughout Africa and is also present in many languages in North and South America, Asia, and Oceania. Many of these are pronounced even when nominal or adjectival predication occurs unembedded (Stassen

1997: 77–91). If nominal and adjectival phrases are of different semantic type, this is certainly not unexpected.

The evidence for the claim that unlike nouns and adjectives, verbs are lexical predicates is substantial. But it is important to keep in mind that this thesis does not automatically yield a semantic characterization of verbs. *Predicate* is the syntactic category of expressions that have a number of arguments whose joint saturation yields a clause. If we assume that all predicates have the same sort of semantic value—for example that they all denote functions to truth-values—we have something that all verbs have in common semantically. Bowers 1993 does make this assumption: he follows Chierchia and Turner 1988 in interpreting PR as a type-shifter mapping properties (conceived of as individuals) to propositional functions.

But the assumption is far from innocent. Raising verbs (for example *seem*) and weather verbs (e.g. *rain*) take expletive subjects as arguments. If we commit ourselves to the idea that these denote functions we have to say what their domains are, which in turn requires assignment of semantic value to the expletive. Perhaps when it seems to Jill that Jack left then *that he left* seems to Jill and perhaps when it rains then *the rain* rains, but these proposals are rather fanciful. The typical assumption is that expletives are semantically vacuous. Based on these sorts of considerations one may be tempted to draw the conclusion that "being a verb is fundamentally a syntactic matter" (Baker 2003: 27)—*seem* and *rain* require a syntactic argument but not a semantic one. In the next section we will see whether the conclusion can be resisted.

2 Verbs Denote Properties of Events

Since the seminal work of Davidson 1967 many have argued that at least some verbs are predicates of events. Motivated primarily by the inferential behavior of manner adverbs, for (4) Davison proposed the logical form (4'):

(4) Jill greeted Jack.
(4') $\exists$e. greeted (Jill, Jack, e)

Followers of Davidson have extended the proposal in two ways. First, they have argued that it would be implausible to think that action sentences differ fundamentally from the rest. A sentence like (5) can also been seen as expressing quantification over underlying entities, except that those entities are states, rather than events.

(5) Jill slept at home.
(5') $\exists$s (slept(Jill, s)$\wedge$ at home(s))

Bach 1986 introduced the term *eventuality* for events, processes, and states. To avoid the slightly awkward term semanticists tend to use the term *event* in such extended sense. I will follow this practice. Second, neo-Davidsoneans have emphasized that the arguments of the three-place predicate in (6) are not on a par. *Jill* and *Jack* are assigned appropriate thematic roles by the verb but the event-argument is non-thematic. We can think of the thematic roles as specifying ways in which Jill and Jack participate in the underlying event—the former as its Agent, the latter as its Theme.

(4") $\exists$e (greet(e)$\wedge$ Agent(e, Jill) $\wedge$Theme(e, Jack)$\wedge$ Past(e))

The extensions of the original Davidsonean proposal have real advantages. In Montague semantics verbs are scattered all over the type-theoretic hierarchy: *play* would be of type <e, t>, *write* of type <e,<e,t>>, *give* of type <e,<e<e,<e, t>>>, etc. In a neo-Davidsonean semantics all verbs have the same type: They are one-place predicates of events. Of course, verbs can be subcategorized by the number of arguments they take. But the linguistically natural way to subcategorize them is more fine-grained than that. For example, among intransitive verbs one may distinguish between *unergative* verbs (for example *run*, *speak*, *laugh*) whose subject is external to the VP and *unaccusative* verbs (e.g. *arrive*, *collapse*, *rust*) whose subject is not. This syntactic difference corresponds to a difference in the sorts of thematic roles assigned. For example the Agent role is always assigned to an external argument while the Theme role is always internal (Gruber 1965, Chomsky 1981, Grimshaw 1990). The distinction is captured nicely in a theory that incorporates thematic relations into logical form.

The neo-Davidsonean approach draws clear distinctions among different aspects of verb meaning. First, there is the core content of the verb, which characterizes an underlying event. In the case of (4"), this is the predicate *greet*, true of all greeting events and nothing else. Second, there is the argument structure of the verb, which determines the number and type of participants in the event. In (4"), this is represented by the second and third conjuncts. Finally, there is the content contributed by the inflection on the verb. Here this is the past tense morpheme represented in (4") by the last conjunct. Separating the three components has been the chief reason Neo-Davidsonean theories have been successful in dealing with difficult phenomena, such as perception reports, aspect, plurals, causatives, focus, secondary predication, and qualification (Higginbotham 1983, Parsons 1990, Schein 1993, Pietroski 2000, Herburger 2000, Rothstein 2002, Szabó 2003, respectively).

Once the different ways verbs contribute to logical form are distinguished, it is possible to capture cases where one or another of these three aspects of verb meaning is vacuous or missing. Thus, representing nonfinite occurrences of verbs is unproblematic:

(6) Jill saw Jack run.
(6') ∃e∃e'(see(e)∧Experiencer(e, Jill)∧Theme(e, e')∧ run(e)∧ Agent(e', Jack)∧ Past (e))

(7) Jack went to run.
(7') ∃e∃e'(go(e)∧ Agent(e, Jack)∧Goal(e, e')∧run(e')∧Agent(e', Jack)∧ Past (e))

(I am ignoring the complication that (7) does not entail that Jack actually ran. Note that the position occupied by *run* is not intensional in the usual sense: one can substitute an extensionally equivalent predicate in this position *salva veritate*.) In (6) and (7) the verbs in the main clauses bear tense marking but the verbs in the embedded clauses do not, and so the seeing and going events are in the past, but the time of the running events is unspecified. This is how the sentences manage to leave it open whether the events they mention are simultaneous. Light verbs pose no problem either:

(8) Jill took a nap.
(8') ∃e (Agent(e, Jill)∧Theme(e, a nap)∧Past(e))

(9) Jill had a nap.
(9') ∃e (Experiencer(e, Jill)∧Theme(e, a nap)∧Past(e))

Take and *have* do not contribute a core component to the logical form of (8) and (9): All these sentences say is that there is a past event whose Agent or Experiencer is Jill and whose Theme is a nap. There is no need to think that the event is some sort of taking or having. Presumably, it is a nap. (The semantics leaves it open whether (8) and (9) describe the *same* events.)

In the same vein, if—as seems plausible—*rain* or *seem* fail to assign thematic roles to their subjects, the logical form of (10) and (11) can simply be as follows:

(10) It rained.
(10') ∃e (raining (e)∧ Past(e))

(11) It seems to Jack that Jill left.
(11') ∃e (seeming (e)∧ Experiencer(e, Jack)∧Theme(e, that Jill left)∧Present(e))

This removes the obstacle for a purely semantic characterization of verbs mentioned at the end of the previous section. All verbs—including weather verbs and raising verbs—are one-place predicates of events. What verbs denote then depends on the semantics of predication one adopts. The simplest view is that verbs plurally denote events, for example that *run* denotes each running event and *sleep* denotes each state of sleeping. Light verbs would denote each event and would be semantically distinguished from one another only in the thematic roles they assign to their arguments. Many light verbs have homonyms that are regular correlates, which would of course denote events within a restricted class. So, *take* in *Jack took the letter from the desk* denotes each taking event and *have* in *Jill has a coat* denotes each state of having. Those who are less concerned about positing entities in the semantics can assign a single type of denotation to all verbs. This could be a function from events to truth-values, a set of events, a kind of events, or a property of events. For convenience, I will adopt the last alternative.

Just because verbs describe events we cannot conclude that other words don't do the same. We certainly seem to be able to talk about events using nouns like *kick*, *freeze*, or *war*. The suggestion that certain entities can only be denoted through certain types of expressions isn't well-motivated. Frege thought predicate denotations (he called them *concepts*) cannot be denoted by nonpredicates, which led him to the conclusion that the concept *horse* is not a concept (Frege 1892: 182–5). It is best to avoid such paradoxes by acknowledging that nominal expressions are referentially unconstrained: anything can (in principle) be referred to by some suitable noun in some suitable language.

What then distinguishes semantically a verb from the corresponding gerund? I suggest that the distinction lies not in *what* they denote but in *how* they denote (Furth 1968, Burge 2007). Verbs, being predicates, are predicated of things; nouns, being referring expressions, refer to them. What *greeting* refers to are greetings and they are precisely what *greet* is predicated of. This does not quite make *greeting* and *greet* synonyms—there is more to meaning than denotation. But *greet* and $[_{\text{PREDP}}\ [_{\text{PR}}\ \varnothing]\ [_{\text{NP}}\ \textit{greeting}]]$ probably do mean the same. Thus the sentence *Jill greeted Jack* can be correctly paraphrased as *some past greeting of Jack is by Jill*.

Functionalist approaches to lexical categories postulate questionable ambiguities (Baker 2003: 15–16). The sentences *Chris hungers*, *Chris has hunger*, and *Chris is hungry* are near synonyms, which is not what one would expect if verbs, nouns, and adjectives denote different sorts of things. The present proposal maintains the functionalist idea that verbs have a distinctive sort of denotation but denies that this is true of nouns and

adjectives. Assuming *has* and *is* are auxiliaries without a core lexical meaning, all three of Baker's sentences describe the very same states—states of hunger Chris is in.

3 Verbs and Times

The definition of a verb advocated here establishes no link between verbs and time. On the one hand, this is a good thing. There are languages where tense can associate with nouns and adjectives (e.g. Turkish, Halkomelem) and languages where it cannot associate with verbs (e.g. Yoruba, Nupe). On the other hand, these languages are the exception (Stassen 1997). Moreover, the auxiliaries often used to express temporal reference (such as the English *will*) typically modify a verb. So while temporality is not essential to being a verb, the correlation is clearly more than accidental.

Givón 1984 argues that verbs typically denote transitory states of affairs, whereas nouns and adjectives pick out more permanent ones. How one locates transitory states of affairs in the flow of time makes a difference, which in turn explains the association of verbs and times. The obvious problem with this proposal is that verbs correlate with transience much less than they do with temporality in general. In addition, the explanation presupposes that *all* lexical categories denote states of affairs—a rather tendentious claim. *Prima facie*, neither *blue* nor *dog* picks out a state of affairs.

According to the event-based view advocated here, the association of verbs with times is less direct than on Givón's account. Verbs denote properties of events and events are temporally structured. Thus the lexical meaning of a verb already contains some temporal information. This is what makes it the natural locus for temporal reference in general. Nouns (and perhaps other words as well) can denote events and their properties, but they don't have to.

Lexical aspect (or *Aktionsart*) is a property of verbs whereby they represent the way the eventualities they describe unfold in time. Lexical aspect gives rise to a natural and cross-linguistically valid classification of verbs; the classic proposal about this goes back to Vendler 1957. He distinguishes four categories: *states*, *activities*, *achievements* and *accomplishments*. States (for example *knowing*) are extended eventualities whose temporal parts could be (and are typically seen as) homogeneous. Activities (for example *walking*) are also temporally extended, but they must have temporal parts distinguishable from one another. Unlike states and activities, achievements (for example *finding*) can be instantaneous and are seen as culminations of processes. Accomplishments (for example *building*) are complex events made up of an activity leading to some resultant state. Vendler presented a number of linguistic tests that classify verbs depending on which of these event types they are true of; Rothstein 2004 has updated the tests and proposed to organize verbs according to the distribution of two lexical features. Some eventualities are *telic* (in the sense of having a natural culmination), others are not; some eventualities have *stages* (in the sense of having discernible temporal parts), others do not:

Table 3.2.1

Eventualities	*+Stage*	*–Stage*
+Telic	accomplishment	achievement
–Telic	activity	state

While there is no consensus about the types of lexical aspect (for a different, equally influential classification see Steedman 1997), it does seem that verbs and only verbs are subject to this sort of classification. This fact is explained by the view that verbs obligatorily denote events, and that we naturally see events as temporally structured. This tendency also explains why verbs are natural loci for temporal reference.

4 Verbs and Statements

The idea that verbs signify affirmation has a long history. Here is how the *Port Royal Logic* sums up the view and its prime motivation (Arnauld and Nicole 1662: 81):

> Why is *Petrus vivit* [Peter lives] a proposition, and *Petrus vivens* [Peter living] not one if you do not add "is" to it as in *Petrus est vivens* [Peter is living]? Only because the affirmation contained in *vivit* was removed from it to make it the participle *vivens*. From this it appears that the affirmation that does or does not exist in a word is what makes it a verb or not.

The obvious objection is that embedded clauses contain verbs but are often not affirmed. The point did not entirely escape the attention of the authors of the *Port-Royal Logic*; they say that in *All philosophers assure us that heavy things fall to earth of their own accord,* the speaker's affirmation is signified only by the verb in the main clause (Arnauld and Nicole 1662: 95). One might hope that discussion of disjunction and conditionals clears up what Arnauld and Nicole think about the signification of embedded verbs. Alas, it does not. Here one could say that the verb of the subordinate clause signifies the affirmation of all philosophers, but this won't help the view in general. The embedded verb *fly* in *Nobody assures us that heavy objects fly away of their own accord* certainly does not signify anybody's affirmation.

Since Frege, we have learned to distinguish between the *force* of an utterance from the *meaning* of the expression uttered. The suggestion that some element within a sentence encodes the speech act the sentence is used to perform ignores this distinction. In principle we could perform all sorts of speech-acts uttering the very same sentence. And one could even perform illocutionary acts uttering nonsentences (Stainton 2005). Just imagine a case when it is in dispute whether Peter is alive and suddenly he enters the room and one utters *Peter living* pointing at him. It is a bit quaint, but it does seem like a *bona fide* statement. So, perhaps we should not say that verbs signify affirmation, but rather that they signify what makes a linguistic expression capable of being true or false. The idea is that without a verb, nothing can express a proposition. This was Russell's view in the *Principles of Mathematics* (Russell 1903: 50):

> A proposition, in fact, is essentially a unity, and when analysis has destroyed the unity, no enumeration of constituents will restore the proposition. The verb, when used as a verb, embodies the unity of the proposition, and is thus distinguishable from the verb considered as a term, though I do not know how to give a clear account of the precise nature of the distinction.

The claim is that *difference* is a "verb considered as a term," not a "verb used as a verb," and this is why the constituents of the proposition *A differs from B*—which would be A, difference, and B—placed side by side do not reconstitute the proposition. But the real

issue is not unity—after all *the difference of A from B* is not a mere list, but an expression exhibiting genuine syntactic unity. It's the kind of unity that matters: without a verb, Russell contends, we cannot get anything that *says* something. Alas, some of the examples mentioned before create a problem for this claim. The embedded clause in *I consider Jill happy* manages to be truth-evaluable without containing a verb. (If you think that without time determination no proposition can be expressed, take the embedded clause of *I consider Jill happy now*.) The embedded clause contains the predicate [$_{\text{PREDP}}$ $^{\varnothing}$ [$_{\text{AP}}$ *happy*]], and this appears to be enough for being true or false. (If you think that *I consider Jill happy* is but a truncated form of *I consider Jill to be happy*, consider instead *I made Jill happy*, which is exceedingly unlikely to be the truncated form of *I made Jill to be happy*.)

What distinguishes the verb *live* from the participle *living* and the verb *differ* from the noun *difference* is that verbs are predicates while participles and nouns are not. The syntactic difference presumably corresponds to a semantic one. As Russell correctly suggests, the difference should not be thought of as a difference in denotation, but rather as a difference in the manner in which these words denote. Predicates say something of what they denote, other expressions don't.

It is not clear why expression of a proposition requires a predicate—to say that only predicates can *say* something is not much more than a restatement of the problem. It does seem that there is a deep connection between truth and predication, one that any serious philosophy of language must ultimately grapple with. But the connection of verbs to this perplexing issue is at best indirect—verbs are the only lexical predicates in natural languages, so the syntactically simplest sentences in these languages can contain predicates only if they contain verbs.

References

Arnauld, A. and P. Nicole (1662) *Logic, or the Art of Thinking*, Cambridge: Cambridge University Press, 1996.

Bach, E. (1986) "The Algebra of Events," *Linguistics and Philosophy* 9: 5–16.

Baker, M. (2003) *Lexical Categories*, Cambridge: Cambridge University Press.

Bar-Hillel, Y. (1953) "A Quasi-Arithmetical Notation for Syntactic Description," *Language* 29: 47–58.

Bowers, J. (1993) "The Syntax of Predication," *Linguistic Inquiry* 24: 591–656.

Bowers, J. (2001) "Predication," in M. Baltin and C. Collins eds., *The Handbook of Contemporary Syntactic Theory*, Oxford: Blackwell.

Burge, T. (2007) "Predication and Truth," *Journal of Philosophy* 104: 580–608.

Chierchia, G. and R. Turner (1988) "Semantics and Property Theory," *Linguistics and Philosophy* 11: 261–302.

Chomsky (1970) "Remarks on Nominalization", in R. Jacobs and P. Rosenbaum eds., *Readings in English Transformational Grammar*, Waltham, MA: Ginn.

Chomsky, N. (1981) *Lectures on Government and Binding*, Dordrecht: Foris.

Davidson, D. (1967) "The Logical Form of Action Sentences," reprinted in *Essays on Actions and Events*, Oxford: Oxford University Press, 2001.

Frege, G. (1892) "On Concept and Object," reprinted in B. McGuinness ed., *Collected Papers on Mathematics, Logic, and Philosophy*, Oxford: Basil Blackwell, 1984.

Furth, M. (1968) "Two Types of Denotation," in *Studies in Logical Theory*, Cambridge: Blackwell.

Givón, T. (1984) *Syntax: A Functional-Typological Introduction*, Amsterdam: John Benjamins.

Grimshaw, J. (1990) *Argument Structure*, Cambridge, MA: MIT Press.

Gruber, J. (1965) Studies in Lexical Relations, PhD Dissertation, MIT.

Herburger, E. (2000) *What Counts: Focus and Quantification*, Cambridge, MA: MIT Press.

Higginbotham, J. (1983) "The Logic of Perceptual Reports: An Extensional Alternative to Situation Semantics," *Linguistics and Philosophy* 80**:** 100–27.

Jackendoff, R. (1977) *X-bar Syntax*, Cambridge, MA: MIT Press.
Parsons, T. (1990) *Events in the Semantics of English*, Cambridge, MA: MIT Press.
Pietroski, P. (2000) *Causing Actions*, Oxford: Oxford University Press.
Rothstein, S. (2001) *Predicates and their Subjects*, Kluwer: Dordrecht.
Rothstein, S. (2004) *Structuring Events: A Study in the Semantics of Lexical Aspect*, Blackwell: Oxford.
Russell, B. (1903) *The Principles of Mathematics*, London: Northon.
Schein, B. (1993) *Plurals and Events*, Cambridge, MA: MIT Press.
Stainton, R. (2005) "In Defense of Non-Sentential Assertion," in Z.G. Szabó ed., *Semantics vs. Pragmatics*, Oxford: Oxford University Press.
Stassen, L. (1997) *Intransitive Predication*, Oxford: Oxford University Press.
Steedman, M. (1997) "Temporality," in J. van Benthem and A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier: North Holland.
Szabó, Z.G. (2003) "On Qualification," *Philosophical Perspectives* 17**:** 385–414.
Vendler, Z. (1957) "Verbs and Times," *Philosophical Review* 66**:** 143–60.

3.3
ADJECTIVES

Chris Kennedy

1 Introduction

In *The Cambridge Grammar of the English Language*, adjectives are characterized as expressions "that alter, clarify, or adjust the meaning contributions of nouns," in order to allow for the expression of "finer gradations of meaning" than are possible through the use of nouns alone (Huddleston and Pullum 2002, p. 526). At a general level, adjectives gain this capability in virtue of two main characteristics, one of which is semantic and one of which is syntactic. On the semantic side, they introduce properties. (Whether they actually denote properties is a question we will address in detail below.) On the syntactic side, they are able to function as modifiers, and so may (with some restrictions) combine recursively with nouns. The result of this combination is a new property which is typically (though not always) true of a subset of the entities that the original properties are true of, thereby providing a "finer gradation of meaning" than is possible using the noun alone. This simple picture hides many important and interesting complexities, however, which provide insights on several topics of central interest to both linguists and philosophers, including: vagueness, contextualism, relativism, compositionality, and the semantic analysis of significant phenomena such as modality. I begin with an examination of the distributional properties of adjectives, then summarize the most prominent analyses of their meanings, and finally conclude with a look at some of the roles that adjectives have played in reasoning about the issues and phenomena mentioned above.

2 Distribution

As it turns out, determining exactly what is constitutive of the grammatical category 'adjective' is not entirely straightforward. There are a number of distributional tests that distinguish adjectives from other categories, as we will see below, but it is not the case that all terms that are traditionally classified as adjectives in a particular language satisfy all of these tests, and it is likewise not the case that the tests apply uniformly across languages to terms that otherwise share the semantic properties that are traditionally thought to be associated with adjectives. To keep things simple, I will focus primarily in this section on adjectives in English, with a few comments here and there about the behavior of adjectives in other languages. The reader should keep keep in mind, however, that although all languages have terms that share the semantic properties of English adjectives, the distributional patterns of these terms can vary. (See Dixon and Aikenvald 2004 for a detailed discussion of the cross-linguisic properties of adjectives.)

The first identifying feature of adjectives involves their use as predicate terms. Like verbs, adjectives may supply the main predicate term in a sentence, and may even introduce their own arguments, as shown by examples such as (1) and (2). (I'll assume here that the verb *be* in (1a) and (2a) is just providing a host for tense and agreement information, and is not playing a central role in the meaning of the predicate. Many languages do not require expression of this element in sentences like these.)

(1) a. That stone is weighty.
    b. That stone weighs a lot.
(2) a. The country is dependent on foreign oil.
    b. The country depends on foreign oil.

However, only adjectives can serve as the complements of the epistemic verbs *seem* and *appear*, as shown by the following contrasts (* denotes syntactic illformedness):

(3) a. That stone seems/appears weighty.
    b. * That stone seems/appears weigh a lot.
(4) a. The country seems/appears dependent on foreign oil.
    b. * The country seems/appears depend on foreign oil.

This test doesn't uniquely pick out adjectives, however: nouns (or rather noun phrases) can sometimes appear as the complement of *seem* and *appear*, especially when their meanings are in some sense scalar or evaluative. This is illustrated nicely in the following lines from *The Ship of Fools* by Sebastian Brandt (which appear on p. 294 of the 1962 edition of Edwin Zeydel's 1944 translation, published by Dover):

(5) He seems a burden, seems a pest
    To all his brood, a hateful guest,
    And yet it almost serves him right,
    For he's a dull and witless wight.

A second diagnostic, which distinguishes adjectives from both nouns and verbs, is the possibility of direct composition with degree words like *rather, very, too, so, enough, how*. For example, of the related terms *dependent, depend* and *dependence*, only the first can directly combine with the excessive degree marker *too*:

(6) a. The country is too dependent on foreign oil.
    b. * The country too depends on foreign oil.
    c. * The country has too dependence on foreign oil.

Examples (6b–c) can be repaired by first combining *too* with *much* (and in the case of (6b), moving the whole thing to the left of the verb, deriving *depends too much on foreign oil* or *depends on foreign oil too much*), but this only serves to illustrate the point that it is only the adjectival form *dependent* that can directly combine with the degree word. It should be emphasized, though, that adjectives accept composition with degree words only to the extent that they are associated with concepts that are, or can be, thought of as scalar, in a sense to be discussed below.

Perhaps the most central diagnostic for the class of adjectives is the one that is implicit in Huddleston and Pullum's functional/semantic characterization of adjectives as expressions that "alter, clarify, or adjust the meaning contributions of nouns": adjectives can directly compose recursively with nouns, forming more complex constituents, which may then combine with other elements (e.g., a determiner or possessive nominal) to form a noun phrase, as in (7a–c).

(7) a. a blue ball
b. a round blue ball
c. a large round blue ball

Such uses of adjectives are referred to as instances of ATTRIBUTIVE MODIFICATION. In some languages, adjectives may only be used attributively. For example, in the Yanaria language of New Guinea, adjectives may directly combine with nouns, as in (8a), but they may provide the main predicate of a sentence only if they compose first with a nominal element meaning 'thing, matter', as shown in (8b); omission of this element results in ungrammaticality.

(8) a. haga' dote'na
tasty food
'tasty food'
b. ma'i egemo haga-na-e'
this banana tasty-thing-PRED
'This banana is tasty.' (Lit. *This banana is a tasty thing*.)

Even English includes a number of adjectives that have only attributive uses, such as *former*, *mere*, *principal* and *main*:

(9) a. This is our former/principal/main objective.
b. * This objective is former/principal/main.

The existence of expressions like these has led some researchers to hypothesize that the attributive use of adjectives is in some important sense basic, a point to which we will return in detail below. However, like the other tests, this one also has exceptions, though they are few and appear to be systematic. For example, there is a class of adjectives which includes *asleep*, *awake*, *alone* (sometimes called *a*-adjectives, for obvious reasons) which can appear as complements of *seem* and *appear*, but are barred from attributive position:

(10) a. * Kim photographed two asleep/alive polar bears.
b. Kim photographed two sleeping/living polar bears.

There are, in addition, languages which require noun-modifying adjectives to first combine with a predicative element, effectively turning them into relative clauses (and calling into question their status as adjectives to begin with; see Baker 2003).

Cases like these show that the possibility of attributive modification is not a necessary condition for adjective status, but it is generally agreed that it is a sufficient one. Nevetheless, some care must still be taken in applying this test. Nouns may also

combine directly with nouns, as in *eyeball*, *tennis ball*, *home run ball*, or *medicine ball*, but in a way that is different from adjectives in two respects. First, the interpretation of such structures (referred to as NOUN-NOUN COMPOUNDS) is variable and often context dependent: an eyeball is a part of the body that has the shape of a ball; a tennis ball is a ball used for playing tennis; a home run ball is a ball that was hit for a home run (e.g., *Barry Bonds' 756th home run ball was auctioned for $752,467*); a medicine ball could be a ball of medicine, a ball used to deliver medicine, or a piece of gym equipment. Attributive adjective modification, in contrast, gives rise to much more systematic and restricted interpretations, as we will see in detail below.

Second, attributive adjectives are different from nouns in compounding structures in that the former cannot occur outside the latter:

(11) a. a majestic towering home run ball
b. * a majestic home run towering ball
c. * a home run majestic towering ball

In contrast, attributive adjectives can often be reordered without compromising syntactic well-formedness:

(12) a. a majestic towering home run
b. a towering majestic home run

Interestingly, it is not the case that attributive adjective ordering is fully unrestricted. For example, the default order of the adjectives *numerous, inefficient*, and *American* as attributive modifiers is as in (13a); orders in which *numerous* is non–initial are ungrammaticsal (13b–c); and an order in which *American* precedes *inefficient* is acceptable just in case *American* is understood contrastively or in focus. For example, (13d), with stress on *American* (indicated by capitalization), would be acceptable as an answer to the question *Are there a lot of inefficient cars on the road?*

(13) a. There are numerous inefficient American cars on the road.
b. * There are inefficient numerous American cars on the road.
c. * There are inefficient American numerous cars on the road.
d. There are numerous AMERICAN inefficient cars on the road
(but not so many JAPANESE ones).

These ordering restrictions are robust cross-linguistically, holding both in languages like English, where adjectives precede nouns, and in a mirror-image fashion in languages in which nouns precede adjectives, though the underlying reasons for the distribution are not well understood (see Demonte 2008; Svenonius 2008; and Cinque 2010 for recent discussion).

Sometimes multiple orders are possible but result in significant differences of interpretation. For example, *wild Minnesotan rice* denotes quantities of uncultivated or unruly rice, which stands in some relation to Minnesota (most likely it was grown there, though other interpretations are possible), while *Minnesotan wild rice* denotes quantities of *zizania palustris* (which is in fact not a species of rice). The relative order of the adjective and the noun, when two orders are possible, can also affect meaning. Consider, for

example, (14), in which the adjective can either be interpreted nonrestrictively, as in (14a), or restrictively, as in (14b) (Bolinger 1967; Larson and Marušič 2004).

(14) All of his unsuitable remarks will be eliminated from the final text.

a. All of his remarks will be eliminated; they are unsuitable.
b. All (and by implication, only) those of his remarks that are unsuitable will be eliminated.

When the adjective occurs postnominally, however, only the restrictive interpretation is available:

(15) All remarks unsuitable (for publication) will be eliminated from the final text.

The examples in (16), discussed originally by Bolinger (1967) (see also Larson 1998; Cinque 1993, 2010; Demonte 2008; and Morzycki 2008), show a similar sensitivity to the relative order of the noun and the adjective.

(16) a. The visible stars include Capella, Betelgeuse and Sirius.
b. The stars visible include Capella, Betelgeuse and Sirius.

(16a) is truth-conditionally ambiguous: it can be understood as a claim about which stars are visible at the time of utterance, or as a claim about which stars are intrisically visible (e.g., capable of being seen by the naked eye). At noon on a sunny day, (16a) would (normally) be false on the first reading and true on the second. Example (16b), in contrast, is unambiguous: It has only the 'currently visible' reading, and would be false in the sunny day context. This suggests that the ambiguity in (16a) does not reflect an ambiguity in *visible*, but rather has something to do with composition. This conclusion is further strengthened by the contrast between (17a) and (17b).

(17) a. The invisible visible stars include Betelgeuse.
b. ?? The visible invisible stars include Betelgeuse.

(17a) means that Betelgeuse is among the stars which can generally be seen but are currently invisible, which is a perfectly coherent thing to say. Example (17b), on the other hand, sounds a bit odd (indicated by the '??'), because it involves definite reference to a set of stars that are currently visible and intrinsically invisible, which is a combination of properties that is difficult to have, and possibly even contradictory. As pointed out by Larson (1998), these facts show that the 'current' vs. 'intrinsic' distinction is not (or not only) a function of the relative ordering of the adjective and the noun but (also) reflects more subtle facts about adjective–noun composition.

Taken together, examples like these show that subtle differences in structure can affect the truth conditions of sentences with attributive adjectives, a fact that must be kept in mind when constructing arguments—either linguistic or philosophical—based on the interpretations of such constructions. Superficially simple structures sometimes hide an underlying complexity, which must be taken into account by reasoning based on the meanings of those constructions and the elements they contain.

3 Meaning

I said above that adjectives introduce properties. Two kinds of facts suggest that adjectives also denote properties. First, as we have already seen, adjectives may provide the main predicate in a sentence. Second, we often see entailments from the attributive form to the predicative form, as in (18).

(18) a. Cosmo is a hairy brown dog.
b. Cosmo is hairy.
c. Cosmo is brown.

We do not need to know anything about dogs to know that (18a) entails (18b–c). If we later learn that Cosmo is not merely a dog but also a Westminster Kennel Club champion, we may also justifiably conclude from (18a) that he is a hairy brown Westminster Kennel Club champion. If (18a) involves the ascription of three properties to Cosmo (being a dog, being brown and being hairy), these patterns of reasoning follow.

Adjectives that give rise to such reasoning patterns are often referred to as INTERSECTIVE. Not all adjectives are intersective, however, a fact that introduces challenges for the idea that adjectives as a class denote properties. Consider the following examples, from Partee 1995. Knowing that (19a) is true does not justify the conclusion in (19b), because it could be the case that the only respect in which Francis is skillful is in his role as a surgeon, in which case we would accept the former but most likely deny the latter.

(19) a. Francis is a skillful surgeon.
b. Francis is skillful.

Similarly, the combined truth of (19a) and (20a) do not license the conclusion in (20b): Francis could be a very skillful surgeon and still have only limited facility with the violin.

(20) a. Francis is a violinist.
b. Francis is a skillful violinist

Intuitively, *skillful* in (19a) and (20b) picks out just the subset of surgeons and violinists who are skillful *as surgeons* and *as violinists* respectively. As a result, we cannot conclude from the truth of e.g. (19a) that Francis is skillful in any other way. Partee (1995) labels adjectives like *skillful* SUBSECTIVE, since composition of the adjective with a nominal constituent returns a subset of the denotation of the nominal, but in a way that does not support the same inferences as with intersective adjectives: All intersective adjectives are subsective, but not all subsective adjectives are intersective.

In addition to intersective and subsective adjectives, there is a third group of what Partee calls NONSUBSECTIVE adjectives including *former*, *alleged*, *fake*, *possible*, *ersatz* and so forth, which are neither intersective nor subsective. The set of objects that satisfy the description *former president of the United States* is neither the intersection of the set of former things (if that even makes sense) with the set of presidents, nor is it a subset of the set of presidents. Similarly, a *fake identification* is arguably not an identification at all. (Nonsubsective adjectives like *fake* which imply exclusion from the noun meaning are sometimes called PRIVATIVE.)

A number of researchers, including Lewis (1970), Wheeler (1972), Cresswell (1973) and Montague (1974), have taken the existence of non-intersective interpretations of adjectives as evidence that adjectives do not denote properties, but rather must be analyzed as expressions that map properties into new properties. (Others have adopted a more nuanced view whereby attributive uses involve such a meaning, while predicative uses denote properties; see Siegel 1976 for a sophisticated implementation of this kind of account, and the kind of linguistic data that can be brought to bear to support it). In some cases (the intersective adjectives), the output is just the conjunction of the input with a property introduced by the adjective; in others (the non-intersective ones), the adjective determines the output property in a more complex way, as we have seen. Furthermore, apparent predicative uses of adjectives are analyzed as deriving from an underlying attributive source, so that what is predicated of the subject in e.g. (19b) is not the property of being skillful, but rather the property of being *a skillful one*, where the value of the anaphor *one* is filled in contextually. English is thus analyzed on a par with languages like Yagaria (see (8b) above), the only difference being that the surface syntax obscures rather than reflects the the underlying form.

The attributive analysis of adjectives represents a kind of "generalization to the worst case" strategy, which can be found elsewhere in compositional analyses of English (cf. the analysis of proper names as generalized quantifiers in Montague 1974). The advantage of such an approach is that it allows for a general theory of lexical types and compositional operations, and if the general goal is to show that the semantic properties of natural language can be accounted for within a compositional framework, the strategy is a reasonable one. (Though see Kamp 1975 for a critical assessement of the explanatory power of the attributive analysis.) The disadvantage of such an approach is that in effectively building noun-dependency into the meaning of the adjective, the uniformity hypothesis doesn't leave much space for complex structural effects on meaning of the sort we observed in the previous section for adjectives like *visible*. At the same time, it can lead to an over-simplistic assessment of the data, when a more sophisticated analysis of both noun and adjective meaning can provide us with ways of explaining patterns like those above without adopting the attributive analysis of adjective meaning.

Consider, for example, (21a), which is ambiguous between the reading in (21b), in which the adjective is subsective, and the one in (21c), in which the adjective is intersective.

(21) a. Lee is a beautiful singer.
b. Lee sings beautifully.
c. Lee is a singer who is beautiful.

Larson (1998) shows that this ambiguity can be captured straightforwardly without positing an attributive semantics for adjectives—i.e., by maintaining the hypothesis that adjectives denote properties—by extending Davidson's (1967) well-established analysis of adverbial modification in action sentences to adjectival modification. Specifically, nouns like *singer* are analyazed as relations between events and individuals, and adjectives are analyzed as properties of either events or individuals. In some cases, the syntax of the surface form determines exactly what kind of argument the adjective has: in (21b) it is an event, and in (21c) it is an individual. (The addition of the suffix *-ly* is also syntactically conditioned.) In others, such as (21a), the syntax is compatible with either option, resulting in ambiguity. Depending on which option we choose for the adjective, we

derive the truth conditions in (22a–b) for the sentence which correspond to the readings in (21b–c), respectively.

(22) a. ∃e[**beautiful**(e) ^ **singer**(**lee**,e)]
b. ∃e[**beautuful**(**lee**) ^ **singer**(**lee**,e)]

Larson's analysis shows how one kind of subsectivity can be handled by appealing to a more sophisticated theory of noun meaning; a different kind of subsectivity, manifested by adjectives that encode scalar concepts, referred to as GRADABLE ADJECTIVES, can be handled by appealing to a more complex view of adjective meaning. Consider, for example, the adjective *tall*. The truth of (23a) does not guarantee the truth of (23b); likewise, knowing that (23a) is true and knowing that Julian is a basketball player does not allow us to conclude (23c).

(23) a. Julian is a tall jockey.
b. Julian is tall.
c. Julian is a tall basketball player.

Facts like these lead to the conclusion that *tall* does not denote a property on its own; instead, *tall* comes to denote a property only after determing a "threshold" or STANDARD of height that an object must reach in order to count as tall, which is itself computed on the basis of a relevant set of objects, or COMPARISON CLASS (Kamp 1975; Klein 1980; Kennedy 2007). To say that Julian is a tall jockey, in other words, is to say that he is a jockey who is tall relative to the standards for jockeys, which does not entail that he is tall relative to some other standard or comparison class. In particular, this does not entail that he is tall relative to whatever standard is appropriate for basketball players, so we are not licensed to conclude (23c), even if we know that Julian is a basketball player.

One way of accounting for facts like those in (23) is to adopt an attributive analysis of adjectives, whereby the noun provides the comparison class for the adjective. On this view, *tall* denotes a function from properties to properties of the form in (24), where **stnd** picks out an appropriate value from the set of heights we get by applying the **height** function to the objects in the denotation of the noun (cf. Klein 1980; Heim and Kratzer 1998; a proper version of this analysis would need to intensionalize the comparison class, of course).

(24) $\lambda P \lambda x.\mathbf{height}(x) \geq \mathbf{stnd}\{\mathbf{height}(y) \mid P(y)\}$

However, a closer look at the distributional properties of adjectives like *tall* shows that this is not the only possible analysis of their meanings, or of facts like those in (23). As the following examples show, a central characteristic of gradable adjectives is that they can appear in a variety of constructions and are linked semantically by encoding different notions of degree: relations to measures, comparison relations, relations of sufficiency and excess, and so forth.

(25) a. Julian is four feet tall.
b. Julian is taller than Sterling.
c. Julian is as tall as we expected him to be.

d. Julian is too tall to fit in the box.
e. Julian is tall enough to reach the ceiling.
f. Julian is so tall that he has to buy special clothes.

These kinds of facts have led many researchers to hypothesize that gradable adjectives do not express relations between properties (or whatever the proper semantic conception of noun meanings is) and properties, but rather relations between more abstract representations of measurement, or DEGREES, and properties (see e.g. Bartsch and Vennemann 1973; Seuren 1973; Cresswell 1976; von Stechow 1984; see Klein 1991 and Bale 2009 for detailed discussions of how degrees can be related to equivalence classes of individuals). The most common implementation of this view posits the denotation in (26) for *tall*, where d is a degree of height.

(26) $\lambda d \lambda x.\mathbf{height}(x) \geq d$

On this view, the function of the complex constituents that combine with the adjective in the examples in (25) is to fix the value of the degree argument, thereby providing a standard of comparison, and turning the adjective into a property that holds of an object if its height exceeds the relevant standard. The predicate in (25a), for example, denotes the property of having a height that exceeds the degree denoted by four feet; the comparative construction in (25b) denotes the property of having a height that exceeds the degree of Sterling's height; and so on. Importantly, the complex constructions denote properties, and indeed behave intersectively in attributive position: (27a) entails (27b), and if Sterling and Julian are basketball players as well as jockeys, we may also draw the conclusion in (27c). (Prenominal comparatives also presuppose that the nominal predicate applies to both the target and standard of comparison, but this is an independent fact which follows from the syntax and compositional semantics of comparatives; see Bresnan 1973.)

(27) a. Julian is a taller jockey than Sterling.
b. Julian is taller than Sterling.
c. Julian is a taller basketball player than Sterling.

Paradoxically, in this kind of analysis, it is unmodified occurrences of gradable adjectives of the sort seen in (23a–c) that present the trickiest analytical challenge, since there is no constituent in the surface form to saturate their degree arguments and turn them into properties. The usual approach is to hypothesize a phonologically null, "positive degree" morpheme which does this job, by existentially binding the degree argument and imposing the restriction that it come from a degree on the scale above a certain threshold. How exactly this threshold is identified, and the extent to which it is determined based on discourse context and linguistic context (e.g., by the fact that an adjective is used attributively vs. predicatively, or by lexical semantic properties related to the kind of scale the adjective uses), are issues that are resolved differently in different analyses (see Kennedy 2007 for a comparison of approaches). The end result is that nonintersectivity in examples like (23a–c) is accounted for not by hypothesizing a function–argument relation between the adjective and the noun, but by hypothesizing a more complex semantic analysis of the adjective (phrase).

That said, it should be acknowledged that, like the attributive analysis, this approach also gives up on the idea that adjectives (the gradable ones, at least) denote properties.

Instead, gradable adjectives denote relations between individuals and scalar values (degrees), and come to denote properties only through composition with something that saturates their degree arguments. This has the analytical advantage of providing a ready account of complex constructions like those in (27) (though such constructions have also been analyzed in nondegree analyses, which begin from semantic assumptions about adjective meaning on a par with (24); see below for details). It also has certain theoretical advantages in the analysis of phenomena of interest to philosophers, such as vagueness and context-dependence, a point I return to below. However, it has a significant disadvantage from a purely linguistic perspective: if gradable adjectives do not directly denote properties but come to do so only through composition with special degree-saturating morphology, then why is it the case that in all the languages of the world that we know of, the linguistic form that we think of as introducing the "core" property associated with the word—the property of being tall, large, rich, happy, and so forth—is never marked by overt morphology? This is a question which so far has not been given a satisfactory answer by proponents of degree-based semantic analyses of gradable adjectives.

4 Philosophical Interest

Adjectives have played a prominent role in a number of philsophical discussions of aspects of human language, as I will document in this section, but perhaps the most prominent is their role in the characterization and analysis of vagueness (Chapter 4.13). The problem of vagueness is essentially the problem of being unable or unwilling to say of any single point along an ordering generated by the meaning of a particular term whether that point separates the things that the term is true of from the things that it is false of. Vagueness is not a feature of adjectives alone, but adjectives provide a particularly rich empirical ground for investigating it, because so many of them fall into the class of gradable adjectives discussed in the previous section, and so (in their basic, unmodified forms) introduce properties that are true or false of objects depending on their position on a scale.

A central question in work on gradable adjectives and vagueness is whether vagueness is the defining characteristic of the class, with their other significant properties, such as the possibility of forming comparative constructions like those in (25) arising as a result of this feature, or whether vagueness is derived. The first view is seen in the work of Wheeler (1972), Kamp (1975), Klein (1980), van Benthem (1982), and most recently by van Rooij (2011), who provide compositional semantic analyses of various kinds of comparative constructions in terms of an initial analysis of gradable adjectives as vague property terms. This approach has the advantage of explaining the apparent morphological universal mentioned at the end of the previous section: if there is a difference in morphological complexity between the positive and comparative form of an adjective, it is always the latter that is complex. (Though it should be noted that many languages—probably the majority—do not make a morphological distinction between the forms; see Ultan 1972.)

The second view is associated with degree-based analyses of gradable adjectives of the sort discussed in the previous section: Since adjectives do not denote properties at all, but rather relations between individuals and degrees, there is no sense in which the basic meanings of the terms are vague. Instead, vagueness is introduced compositionally through the mapping of such relations to properties. In particular, if this mapping is achieved through composition with a phonologically null "positive" morpheme, as

described above, this opens up the analytical possibility of associating vagueness with the particular semantic features of this morpheme, a move advocated and justified by Fara (2000) and Kennedy (2007, 2011).

Adjectives have also played an important role in discussions of the implications of variable judgments about truth for theories of meaning. Recent work on semantic relativism (see Chapter 4.15) has focused extensively on differences in truth judgments of sentences containing adjectives of personal taste like *tasty* and *fun* (see e.g. Richard 2004; Lasersohn 2005; MacFarlane 2005; Stephenson 2007; Cappelen and Hawthorne 2009), and researchers interested in motivating contextualist semantic analyses have often used facts involving gradable adjectives (recall the judgments in (23) which show that the threshold for what "counts as" tall can change depending on whether we are talking about jockeys or basketball players) to develop arguments about the presence (or absence) of contextual paramters in other types of constructions, such as knowledge statements (see e.g. Unger 1975; Lewis 1979; Cohen 1999; Stanley 2004, and Chapters 3.7 and 4.14). Other researchers have attempted to account for the apparent context sensitivity of these examples without importing context dependence into the semantics (see e.g. Cappelen and Lepore 2005).

More radically, Charles Travis (1997, 1985, 1994) has used judgments about the truth of sentences containing color adjectives to argue against the view that sentences determine truth conditions. Instead, according to him, the semantic value of a sentence at most imposes some necessary conditions under which it may be true (as well as conditions under which it may be used), but those conditions need not be sufficient, and the content of the sentence does not define a function from contexts to truth. However, Travis's argument goes through only if it is the case if the truth conditional variability introduced by color adjectives cannot be linked to a context-dependent element in its logical form or to an underlying ambiguity. And indeed, there are responses to his work which argue for each of these positions based on careful and sophisticated linguistic analysis of color adjectives (see e.g. Szabó 2001; Rothschild and Segal 2009; Kennedy and McNally 2010).

The significance of adjectives for general questions about compositionality in language goes well beyond cases like Travis's. Two additional kinds of phenomena are of particular interest. The first involves sentences like the following, in which the adjective *slow* seems to be contributing a different shade of meaning depending on the sort of thing it is predicated of: a slow quarterback is one who runs (or maybe executes plays) slowly; a slow road is one on which traffic moves slowly; a slow song is one with a slow tempo; and a slow book is one that takes a long time to read.

(28) a. Tom Brady is a slow quarterback.
b. Lake Shore Drive is a slow road during rush hour.
c. "Venus in Furs" is a slow song.
d. *Remembrance of Things Past* is a slow book.

In each case, the meaning contributed by the adjective appears to be systematically related to a kind of activity that is conventionally associated with the meaning of the noun. This has led some researchers to hypothesize that the compositional relation between adjectives and nouns can be even more complex than what we saw above for *beautiful singer*, and in particular that it requires a highly articulated lexical semantic structure for nouns (see e.g. Pustejovsky 1991, 1995; see Fodor and Lepore 1998 for an opposing view).

The second kind of case involves examples in which prenominal adjectives appear to have interpretations outside of the noun phrases in which they appear. For example, adjectives like *occasional, sporadic*, and *rare* can syntactically compose with a noun but have a sentence-level interpretation equivalent to an adverb of quantification, as in the following passage from a 1989 article in the *Chicago Tribune* (where the adverbial occurrence in the third sentence highlights the fact that the adjectival use in the second sentence has a sentence-level meaning):

(29) "I used to be a pretty good Scotch drinker," [Tower] said. "I haven't tasted Scotch in 12 years. After that I had only wine and perhaps an occasional martini, occasionally a little vodka with smoked salmon or caviar or something like that. But that was just occasionally."

Prenominal *average* provides an even more striking example of this kind of phenomenon: (30a) (from a post on answers.bloglines.com) means that the average number of people in an American family is 3.14, and so does not give rise to the bizarre inference that there are actual families which contain 3.14 people, in contrast to (30b), which does give rise to this inference.

(30) a. The average American family consists of 3.14 people.
b. ?? The typical/normal/usual American family consists of 3.14 people.

The fact that (30a) has the meaning that it does indicates that *average* can somehow compose at the sentence level with the numeral and a measure function (which returns the number of whole people in an American family), rather than directly with the property denoted by *American family*. The analytical challenge presented by these cases is to show that the actual meanings can be compositionally derived from the surface forms without resorting to ad hoc stipulations and construction-specific rules; this project is undertaken by Stump (1981); Larson (1998); Gehrke and McNally (2009) for *occasional* and related terms, and by Carlson and Pelletier (2002) and Kennedy and Stanley (2009) for average. While the analyses advocated in these different papers are distinct, they collectively illustrate a theme that is present in much of the work on adjectives described in this chapter: that the semantic properties of adjectival constructions are often more complex than superficial appearances indicate, and must be interpreted against the backdrop of a sophisticated linguistic analysis.

References

Baker, M. C. (2003) *Lexical Categories: Verbs, Nouns, and Adjectives*. Cambridge: Cambridge University Press.

Bale, A. (2009) "A universal scale of comparison". *Linguistics and Philosophy*, 31: 1–55.

Bartsch, R. and Vennemann, T. (1973) *Semantic Structures: A Study in the Relation between Syntax and Semantics*. Frankfurt: Athäenum Verlag.

Bolinger, D. (1967) "Adjectives in English: Attribution and predication". *Lingua*, 18: 1–34.

Bresnan, J. (1973) "Syntax of the comparative clause construction in English". *Linguistic Inquiry*, 4: 275–343.

Cappelen, H. and Hawthorne, J. (2009) *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Cappelen, H. and Lepore, E. (2005) *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.

Carlson, G. and Pelletier, F. J. (2002) "The average American has 2.3 children". *Journal of Semantics*, 19: 73–104.

Cinque, G. (1993) "A null theory of phrase and compound stress". *Linguistic Inquiry*, 24: 239–297.
——(2010) *The Syntax of Adjectives*. Cambridge, MA: MIT Press.
Cohen, S. (1999) "Contextualism, skepticism and the structure of reasons". *Philosophical Perspectives*, 13: 57–89.
Cresswell, M. (1973) *Logics and Languages*. London: Methuen.
Cresswell, M. J. (1976) "The semantics of degree". In B. Partee (ed.), *Montague Grammar*. New York: Academic Press, 261–292.
Davidson, D. (1967) "The logical form of action sentences". In N. Rescher (ed.), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.
Demonte, V. (2008) "Meaning-form correlations and adjective position in Spanish". In L. McNally and C. Kennedy (eds.), *Adjectives and Adverbs: Syntax, Semantics and Discourse*. Oxford: Oxford University Press, 71–100.
Dixon, R. M. W. and Aikenvald, A. Y. (2004) *Adjective Classes: A Cross-Linguistic Typology*. Oxford: Oxford University Press.
Fara, D. G. (2000) "Shifting sands: An interest-relative theory of vagueness". *Philosophical Topics*, 20: 45–81.
Fodor, J. and Lepore, E. (1998) "The emptiness of the lexicon: Reflections on James Pustejovsky's *the generative lexicon*". *Linguistic Inquiry*, 29: 269–288.
Gehrke, B. and McNally, L. (2009) "Frequency adjectives and assertions about event types". In *Proceedings of Semantics and Linguistic Theory 19*. Ithaca, NY: CLC Publications.
Heim, I. and Kratzer, A. (1998) *Semantics in Generative Grammar*. Oxford: Blackwell.
Huddleston, R. and Pullum, G. K. (2002) *The Cambridge Grammar of the English Language*. Cambridge, Cambridge University Press.
Kamp, H. (1975) "Two theories of adjectives". In E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press, 123–155. Reprinted in 1984.
Kennedy, C. (2007) "Vagueness and grammar: The semantics of relative and absolute gradable predicates". *Linguistics and Philosophy*, 30: 1–45.
——(2011) "Vagueness and comparison". In P. Egré and N. Klinedinst (eds.), *Vagueness and Language Use*. Houndmills, UK: Palgrave, 73–97.
Kennedy, C. and McNally, L. (2010) "Color, context and compositionality". *Synthese*, 174: 79–98.
Kennedy, C. and Stanley, J. (2009) "On *average*". *Mind*, 118.
Klein, E. (1980) "A semantics for positive and comparative adjectives". *Linguistics and Philosophy*, 4: 1–45.
——(1991) "Comparatives". In A. von Stechow and D. Wunderlich (eds.), *Semantik: Ein internationales Handuch der zeitgenössischen Forschung* (Semantics: an international handbook of contemporary research), Berlin: de Gruyter, 673–691.
Larson, R. (1998) "Events and modification in nominals". In D. Strolovitch and A. Lawson (eds.), *Proceedings from SALT VIII*. Ithaca, NY: CLC Publications, 145–168.
Larson, R. and Marušič, F. (2004) "On indefinite pronoun structures with APs: Reply to Kishimoto". *Linguistic Inquiry*, 35: 268–287.
Lasersohn, P. (2005) "Context dependence, disagreement, and predicates of personal taste". *Linguistics and Philosophy*, 28: 643–686.
Lewis, D. K. (1970) "General semantics". *Synthese*, 22: 18–67.
——(1979) "Scorekeeping in a language game". *Journal of Philosophical Logic*, 8: 339–359.
MacFarlane, J. (2005) "Making sense of relative truth". *Proceedings of the Aristotelian Society*: 321–339.
Montague, R. (1974) "English as a formal language". In R. Thomason (ed.), *Formal Philosophy: Selected Papers*. New Haven: Yale University Press, 188–221.
Morzycki, M. (2008) "Nonrestrictive modifiers in nonparenthetical positions". In L. McNally and C. Kennedy (eds.), *Adjectives and Adverbs: Syntax, Semantics and Discourse*. Oxford University Press, 101–122.
Partee, B. H. (1995) "Lexical semantics and compositionality". In L. Gleitman and
M. Liberman (eds.), *Language. An Invitation to Cognitive Science*. Cambridge, MA: MIT Press, 311–360.
Pustejovsky, J. (1991) "The generative lexicon". *Computational Linguistics*, 17.
——(1995) *The Generative Lexicon*. Cambridge: MIT Press. Richard, M. (2004) "Contextualism and relativism". *Philosophical Studies*, 119: 215–242.
Rothschild, D. and Segal, G. (2009) "Indexical predicates". *Mind and Language*, 24: 467–493.
Seuren, P. A. (1973) "The comparative". In F. Kiefer and N. Ruwet (eds.), *Generative Grammar in Europe*. Dordrecht: Riedel, 528–564.
Siegel, M. E. (1976) "Capturing the Adjective". Ph.D. thesis, University of Massachusetts at Amherst.
Stanley, J. (2004) "On the linguistic basis for contextualism". *Philosophical Studies*, 119: 119–146.

Stephenson, T. (2007) *Towards a Theory of Subjective Meaning*. Ph.D. thesis, Massachusetts Institute of Technology.
Stump, G. (1981) "The interpretation of frequency adjectives". *Linguistics and Philosophy*, 4: 221–257.
Svenonius, P. (2008) "The position of adjectives and other phrasal modifiers in the decomposition of DP". In L. McNally and C. Kennedy (eds.), *Adjectives and Adverbs: Syntax, Semantics and Discourse*. Oxford: Oxford University Press, 16–42.
Szabó, Z. (2001) "Adjectives in context". In I. Kenesei and R. M. Harnish (eds.), *Perspectives on Semantics, Pragmatics and Discourse: A Festschrift for Ferenc Kiefer*. Amsterdam: John Benjamins, 119–146.
Travis, C. (1985) "On what is strictly speaking true". *Canadian Journal of Philosophy*, 15: 187–229.
——(1994) "On constraints of generality". *Proceedings of the Aristotelian Society*, 94: 165–188.
——(1997) "Pragmatics". In B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell, 87–106.
Ultan, R. (1972) "Some features of basic comparative constructions". In *Working Papers on Language Universals, No. 9*. Stanford, CA: Language Universals Project, Committee on Linguistics, 117–162.
Unger, P. (1975) *Ignorance*. Oxford: Clarendon Press.
van Benthem, J. (1982) "Later than late: On the logical origin of the temporal order". *Pacific Philosophical Quarterly*, 63: 193–203.
van Rooij, R. (2011) "Implicit versus explicit comparatives". In P. Egré and N. Klinedinst (eds.), *Vagueness and Language Use*. Houndmills, UK: Palgrave Macmillan, 51–72.
von Stechow, A. (1984) "Comparing semantic theories of comparison". *Journal of Semantics*, 3: 1–77.
Wheeler, S. (1972) "Attributives and their modifiers". *Noûs*, 6: 310–334.

3.4
QUANTIFIERS AND DETERMINERS

*G. Aldo Antonelli and Robert C. May**

1 Introductory Remarks

When we think about language, and, in particular, about how we use it to speak of things, there is an intuition that permeates virtually all of our thinking and analysis. The intuition is that there is in language an inherent cleavage in the class of terms we use to carry out this task, between those used to speak about specific particulars, and those used to speak in general about those particulars. From the semantic point of view, this is the distinction between reference and quantification; it is what is paradigmatically expressed in natural languages by the distinction within the noun phrases between proper names like "Max" or "Maria" and quantifiers, determiner–noun complexes like "everybody" or "three women." It is not hard to read much, perhaps even most, of the ongoing discussion in the philosophy of language as trying to give content to this intuition; indeed, the field itself dates its birth in part to Russell's reflections on just this point in "On Denoting," when he considers whether definite descriptions are to be analyzed as akin to names or quantifiers, opting for the latter. Russell's arguments, assuming as they do the underlying intuitive validity of the distinction between reference and quantification, however, raise a meta-question: What is one opting for when adopting a quantificational analysis? Or to put it a little differently: What does it mean to say that a term is a quantifier, as opposed to a name? What constitutes being a quantifier? Giving answers to questions of this sort of course was not Russell's primary concern,[1] but it is ours, and so our goal in this chapter is to acquaint the reader with the answers that have emerged, as well as present some novel reflections on these matters.

In searching for answers, a natural first-place to alight is with something elementary. So consider the two first-order quantifiers "there is" and "for all" that are the staples of introductory logic courses, and are employed with insouciance in mathematics, linguistics, and computer science courses throughout the undergraduate curriculum. Typically, these quantifiers are introduced as expressions that can be used in conjunction with a *variable* and an *open formula* to produce a *sentence*. Students are then told that of these quantifiers there are at most two: the existential quantifier $\exists$ and the universal quantifier $\forall$ that combine with variables such as x and y and formulas $\varphi(x)$ or $\varphi(y)$ to produce sentences $\exists x\varphi(x)$ or $\forall y\varphi(y)$, intuitively saying that something (in some intended domain or context) is a φ, or that everything in that same domain is φ. Students are cursorily told that, grammatically, quantifiers are used to bind variables, and after being

shown a few examples of the notation's intended meaning, they are sent on their merry way. In fact, students are seldom, if ever, told *what* the quantifiers are, or for that matter, *why* there are two (or maybe one, given interdefinability with negation, or perhaps more), or *how* they affect the meaning of the expressions in which they occur.

This scenario is perhaps a bit fanciful, but nevertheless it frames questions that are not only substantive, but also foundational; questions which if left unanswered would leave us hard-pressed to say that the intuitive difference between reference and quantification is of any analytic significance. What the answers are, it turns out, are already implicit in our fanciful scenario, at least in broad form. Thus, the very terms in which the scenario is described imply an answer to the first, and most fundamental, question of what quantifiers are. They are *logical* terms, integral aspects of the purely logical machinery; indeed, that it was a logical analysis was a central selling point for Russell's theory of descriptions. Making good on this answer, however, and how it founds answers to the latter two questions, directs us to attend to the manner in which sentences containing quantifiers are represented. Thus, the sentences $\exists x\varphi(x)$ and $\forall y\varphi(y)$ represent three characteristics essential to understanding quantification. The first is *scope*. This is indicated by the use of bracketing conventions, and is essential, as Frege was the first to observe, for properly describing the meaning of sentences in which more than one quantifier occurs, such as $\exists x(\forall y(\psi(x,y)))$. The second is the *meaning* of the quantifier. This is indicated by the symbols "$\forall$" and "$\exists$," whose meanings are standardly glossed as the universal and the existential. The third is the *order* of the quantifier, and this is indicated by the letter following the quantifier symbol. $\exists x\varphi(x)$ contains a first-order existential quantifier, but we can also have quantifiers of higher-orders; thus, $\exists P\, P(a)$ contains a second-order existential quantifier. What we have learned since Frege and Russell's time is that it is through attention to the interaction of these three factors—scope, meaning and order—that there emerges a notion of quantifiers as logical terms that is suffciently articulated to distinguish them from referential terms, and thus ground the *ur*-intuition of the distinction between reference and quantification.

Now, the factors just mentioned conform in an obvious way to a distinction between syntax and semantics. Thus scope is inherently syntactic, a matter of the logical form of sentences containing quantifiers, and, as Tarski showed, if these sentences are composed in the proper manner—that is with the sort of recursive syntax initiated by Frege's "conceptual notation"—then it is possible to construct in parallel a definition of truth suffcient for characterizing logical consequence (Tarski 1956). If this is to be our distinction between syntax and semantics, then an answer to the second of our questions is provided, since for the task at hand we need only one quantifier namely, universal generalization.[2] All we need is to be able to state that matters hold generally, for all things. To be sure, this is a profound result, but it is also limiting in its pragmatism; since just one is needed for the job at hand, the question of how many quantifiers there are is moot. This sort of functional answer will thus leave us short, if our goal is to answer this question in any robust way. Such an answer will require something more than just an account of the logical form of generalization, and the semantics that depends upon it. It will also require an account of the logical *content* of quantification, of the *quantities* that are the content of quantifiers, and not just merely an account of generalization. Giving this sort of robust account is what falls under the purview of the theory of *generalized* quantifiers.

The theory of genralized quantifiers has proven to be of ongoing interest largely because of its application in the semantics of natural language, where its utility has been

central to the analysis of the broad class of determiner expressions, including not only those that correspond to the classical logical expressions, like "every" and "some," but also to words such as "three" and "most" and "only" that do not.[3] The theory has been remarkably frutiful in isolating those properties that unify these words into a class, and distinguishing them from those that differentiate them into natural subgroups. Our goal in this chapter will be to outline answers the theory of generalized quantifiers provides to the questions we posed above, and this will draw our focus largely to how the theory characterizes the meanings of quantifiers. But in doing so, we will also see how the theory articulates the third factor that those answers demand, the *order* of quantifiers. This matter has had a vexed history, primarily because of a confluence of meta-logical and ontological worries that has given rise to an embedded skepticism about second-order logic.[4] Thus, fundamental results obtainable for first-order logic, for example completeness, do not obtain for second-order logic. Moreover, while it is perhaps clear what meets the call of first-order variables, it is by no means as clear at the second-order: Are they to be concepts, properties, attributes, sets? First-order logic, while perhaps acceptable along these grounds, is however too weak to express substantive mathematical results, as opposed to second-order systems that contain terms that generalize over the meanings of predicates. Our novel observations (in section 4) will come on this point; what we will see is that in the context of a robust system of quantification, (not just of generalization)—that is, in generalized quantifier theory—a notion of first-orderizability can be articulated that overcomes the expressive limitations of first-order logic, but without affecting the semantical features associated with the first order. That is, the theory of quantification that analytically grounds our initial intuition regarding the bifurcation of terms, and which affords substantial explanations of the logical and linguistic properties of quantifiers, also founds the notions needed to derive the core propositions of cardinal arithmetic.

In what follows, with our framing questions in mind, we will survey aspects of generalized quantifier theory. There are three parts. In the first, we outline the salient roots of the theory, noting the emergence of the notion of quantifiers as higher-level predicates. In the second, we turn to an overview of contemporary developments directed at articulating set-theoretic foundations for quantifiers, and the relation of first-order definability to first-order semantics as it emerges in this context. We discuss here quantifiers that while not first-order definable do have a first-order semantics, and the importance of these quantifiers for systems that can have significant mathematical content. In the final section, we consider the extension of the theory from standard to nonstandard semantics for first-order quantifiers, and note some of the ramifications of this generalization.

2 The Traditional View

The theory of generalized quantifiers traces its origins all the way back to Aristotle's investigation in the *Prior Analytics* of the logic of the binary determiners Every, Some, No, and Not all. From a grammatical point of view, these quantifiers are expressions taking predicates *S* and *P* as arguments and returning a sentence as a value (sentences obtained in this way are traditionally referred to as "categorical propositions"). The logical relations connecting categorical propositions are codified in the traditional doctrine of the square of opposition (see Parsons 2006) and theory of the syllogism (see Figure 3.4.1).

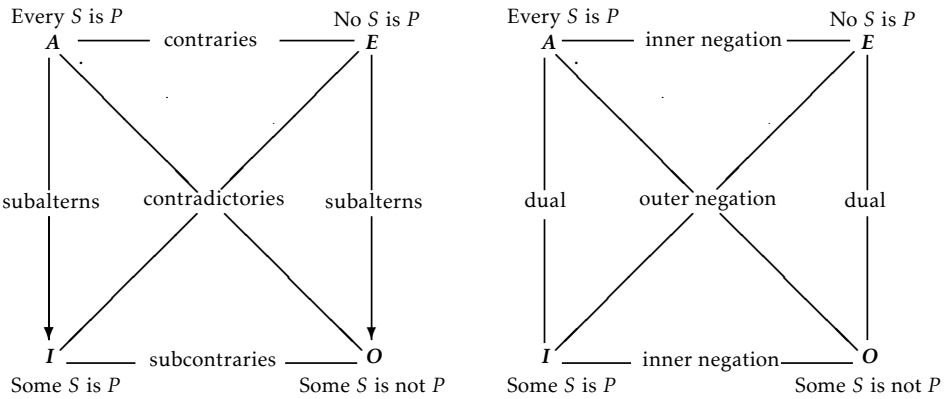


Figure 3.4.1 The traditional square of opposition (left) and the modern one (right).

In the traditional square, if two propositions are *contradictories*, then one of them must be true and the other false; in other words, they can be obtained from each other by applying (outer) negation. So, "Every *S* is *P*" and "Some *S* is not *P*" are contradictories. If a proposition is *subaltern* to another one, then it is implied by it. So "Every *S* is *P*" implies "Some *S* is *P*," and "No *S* is *P*" implies "Some *S* is not *P*." Thus, the two universal forms at the top of the traditional square have *existential import*. If two propositions are *contraries* then they cannot both be true (though they can both be false); this is the case for "Every *S* is *P*" and "No *S* is *P*" (again, because of the existential import of the positive universal). And finally, if two propositions are *subcontraries* then they cannot both be false (though they can both be true), as is the case for "Some *S* is *P*" and "Some *S* is not *P*" (since each is implied by the contradictory of the other one). It should be noted, however, that a case can be made that only the *positive* universal was regarded as having existential import in ancient and medieval times, a view that is perfectly coherent (again, see Parsons 2006 for details).

When the four categorical propositions are translated into modern logical notation, the two universal forms "Every *S* is *P*" and "No *S* is *P*" lose their existential import, with the result that only the contradictory connections along the diagonal survive the transition from the traditional square to the modern one. As mentioned, the contradictory connection can be thought of as "outer negation": "No *S* is *P*" is obtained as the propositional denial of "Some *S* is *P*" (and conversely), and likewise "Some *S* is not *P*" is the propositional denial of "Every *S* is *P*" (and conversely). However, the modern square incorporates also a form of "inner negation," which applies to the predicate term *P*. So the inner negation of "Every *S* is *P*" is "Every *S* is not *P*" (i.e., "No *S* is *P*") and conversely the inner negation of "No *S* is *P*" is "No *S* is not *P*" (i.e.,"Every *S* is *P*"). And in a similar manner "Some *S* is *P*" and "Some *S* is not *P*" are each the inner negation of the other one.

The two kinds of negation, inner and outer, can be combined to obtain the *dual transformation*: The dual of "Every *S* is *P*" is obtained by first taking the inner negation, yielding "No *S* is *P*," and then the outer negation, yielding "Some *S* is *P*." Equivalently, the dual of "Every *S* is *P*" can also be obtained by taking the outer negation first, followed by the inner negation. The dual of "No *S* is *P*" can be obtained in a similar manner by applying the two kinds of negation (in either order). As pointed out by Peters

and Westerståhl (2006, p. 26) any number of applications of the three transformations to any quantifier in the modern square always leads to another quantifier in the square; in other words, it's impossible to land outside of the square by applying the transformations to any of the quantifiers. Moreover, this is not peculiar to the four Aristotelian determiners: *any* quantifier gives rise to its own square of opposition by application of inner negation, outer negation, and dual.[5]

The scientific study of the quantifiers as higher-order operators can be found again at the very beginning of modern logic. In §21 of Frege's *Grundgesetze der Arithmetik* we are asked to consider the forms:

$$\vdash^{\mathfrak{a}}\vdash a^2 = 4 \text{ and } \vdash^{\mathfrak{a}}\vdash a > 0.$$

These forms of the "conceptual notation" correspond to the modern formulas $\exists(a^2 = 4)$ and $\exists a(a > 0)$ (there exists a square root of 4, and there exists a positive number). These forms can be obtained from $\vdash^{\mathfrak{a}}\vdash \varphi(a)$ by replacing the function-name placeholder $\varphi(\xi)$ by names for the functions $\xi^2 = 4$ and $\xi > 0$ (and the placeholder cannot be replaced by names of objects or of functions of 2 or more arguments). These two functions take numbers as arguments and return the value *true* if those numbers are square roots of 2 or (respectively) positive, and *false* otherwise. The kind of arguments of the functions determines their *level*; since these arguments are first-level objects (in this case, numbers) they are referred to as *first-level* functions. First-level functions that return the values *true* and *false* are called *concepts*. So the forms displayed above can be regarded as values of the same function for different arguments. Now, these arguments, as we just noted, are themselves functions. It follows that the form

$$\vdash^{\mathfrak{a}}\vdash \varphi(a)$$

is a *second-level function*, because its arguments are *first-level functions*. Since the arguments of the form $\vdash^{\mathfrak{a}}\vdash \varphi(a)$ are functions that return truth values (i.e., concepts) the same holds for the form $\vdash^{\mathfrak{a}}\vdash \varphi(a)$. Such a form is therefore a *second-level concept*. But such a function is also, as one can easily see, a *quantifier*, for it combines with a variable a and a formula φ to produce a sentence with a determinate truth value, *true* or *false*. This is not a mere coincidence, but a general fact: from a Fregean point of view, quantifiers are second-level concepts.

3 The Modern View

The modern view of quantifiers originating with Frege was further developed under the heading of *generalized quantifiers* by Mostowski (1957) and Montague (1974). This work on generalized quantifiers spans linguistics and mathematical logic, the linguists focusing on quantifiers within natural language semantics, and the logicians focusing on the expressive power and properties such as axiomatizability, decidability etc. The best currently available systematic treatment of these issues is Peters and Westerståhl (2006).

The modern approach takes a characteristically general stance on the matter. By a *domain* (of discourse) we just mean a nonempty set D. A quantifier (viewed as a linguistic—syntactic—entity) can be used in conjunction with a formula to yield a sentence. In turn, a formula is understood semantically as a subset of D, i.e., the collection of those $d \in D$ that satisfy the formula. Accordingly, a quantifier can be identified with a

collection of subsets of D: for instance, the existential quantifier can be identified with the collection of all nonempty subsets of D, and the universal quantifier with the collection containing D itself as its only member. To see how this works, consider that a sentence of the form $\forall x\varphi(x)$ is true over D precisely when every $d \in D$ satisfies φ, i.e., when the extension of $\varphi(x)$ over D is D itself. Hence, $\forall$ can be identified, semantically, with $\{D\}$. Similarly, for the existential quantifier. In general, a quantifier Q over D, then, is just a collection of subsets of D: $Q \subseteq \mathcal{P}(D)$:

1. $\forall = \{D\}$;
2. $\exists = \{X \subseteq D: X \neq \emptyset\}$;
3. $\exists!^k = \{X \subseteq D: |X| = k\}$;
4. $\mathsf{John} = \{X \subseteq D: \text{John} \in X\}$.

The quantifier $\exists!^k$ ("there are exactly k objects such that . . .") denotes the collection of all k-membered subsets of D. The last item above is referred to as a "Montagovian individual": It was Montague who proposed to identify proper names such as "John" with particular quantifiers, i.e., with those subsets of the domain of which John is a member. Then "John runs" is true if and only if the predicate "runs" falls under the quantifier John i.e., if and only if "runs" is among the subsets containing John, i.e., if and only if John runs.

Some quantifiers are best viewed as n-ary *relations* over $\mathcal{P}(D)$. In fact, *binary* relations over $\mathcal{P}(D)$ play a particularly important role, in that they supply denotations for ubiquitous natural language *determiners*. A determiner is an expression such as "every," "no," "most," "between 3 and 5," etc., which combines with a predicate to form a noun phrase. Here are some examples beginning with the Aristotelian determiners:[6]

- Every A is B: $\mathsf{Every} = \{(A,B): A \subseteq B\}$
- Some A is B: $\mathsf{Some} = \{(A,B): A \cap B \neq \emptyset\}$
- Most A are B: $\mathsf{Most} = \{(A,B): |A \cap B| > |A - B|\}$;
- Twice as many A as B are C: $\mathsf{Twice} = \{(A,B,C): |A \cap C| = 2 \cdot |B \cap C|\}$.

It is well known that some binary quantifiers can be represented by means of the corresponding unary quantifier applied to a Boolean combination of their arguments:

$$\mathsf{Every}[A,B] = \forall[-A \cup B],$$

but for instance Most cannot be so represented. In fact, Most is not even *definable* in first-order logic, i.e., there is no formula φ of first-order logic using 1-place predicates A and B as the only nonlogical constants that is true (over some domain D) if and only if (in D) most A's are B's. More generally, we say that a (binary) quantifier $Q(A,B)$ is *first-order definable* over D in a language $\mathcal{L}$ if and only if there is a formula $\varphi \in \mathcal{L}(P,Q)$ such that

$$(D, A,B) \models \varphi,$$

where $\mathcal{L}(P, Q)$ is the language obtained from $\mathcal{L}$ by adding two new predicate symbols P and Q, and (D, A,B) is the $\mathcal{L}$-structure with domain D assigning A and B to P and Q (respectively) as their denotations. For instance, $\mathsf{At\ least\ two}$, as students learn in introductory logic courses, is first-order definable in the pure language of identity.

A quantifier $Q(A,B)$ is *proportional* (over finite D) if and only if there are integers n and m such that $Q(A,B)$ holds precisely when at least of the A's are B's:

$$Q(A,B) \iff \frac{|A \cap B|}{|A|} \geq \frac{m}{n}$$

The quantifiers Most, At least half, More than 10%, . . . are all proportional. Proportional quantifiers are *not* first-order definable.

Quantifiers are distinguished by their *arity* as well as by the number of their *dimension*: The former is the number of arguments (formulas) they take, and the latter is the number of variables of such formulas. ***Every(A,B)*** is a binary monadic quantifier, whereas $Q(R) = 1 \Leftrightarrow \forall x \forall y Rxy$ is a unary dyadic quantifier. In general, the type of a quantifier Q can be represented by a k-tuple $<n_1, \ldots, n_k>$, where k represents the arity, and n_i the dimension of the i-th argument. Formally:

$$Q \in (n_1, \ldots, n_k) \iff Q \subseteq \prod_{i=1}^{k} \mathcal{P}(D^{n_i}).$$

A special case that will be of interest to us is that of *monadic* quantifiers, i.e., a quantifier of type $<1, \ldots, 1>$; if the type specification contains n occurrences of "1", then we have an n-nary monadic quantifier.

The quantifiers $\forall$, $\exists$, $\exists!^k$ are all of type $<1>$; the Aristotelian determiners are of type $<1,1>$, the quantifier Q defined above is of type $<2>$, etc. Since the last one is often called a "dyadic" quantifier, the dimension of a quantifier is sometimes referred to, in analogy with its "arity," as the "adicity."

The distinction between first-and second-order quantifiers can also be made explicit by the assignment of types. Consider the sentence $\exists P \phi(P)$. If we identify the open formula $\phi(P)$ (with P as a free variable) with the class of predicates that satisfy it, then the sentence is true if some P satisfies ϕ, i.e., if the class of predicates satisfying $\phi(P)$ is not empty. It follows that the second-order quantifier $\exists$ denotes the collection $\{X \in \mathcal{P}^2(D): X \neq \emptyset\}$ of all nonempty collections of predicates (here $\mathcal{P}^2(D)$ is the power-set of the power-set of D). Since the class of predicates satisfying $\varphi(P)$ is a first-order quantifier (of type $<1>$ if P is a 1-place predicate), then the second-order quantifier $\exists$ has type $<<1>>$. It's important to stress that *the distinction between first- and second-order quantifiers is semantical, not merely notational.* The distinction also does not coincide with whether a given quantifier is *representable* in first-order logic, as in the case of the quantifier Most.

We take the time to review here several important properties that have been singled out for binary quantifiers $Q(A,B)$:

- Conservativity: $Q(A,B) = Q(A, A \cap B)$;
- Right monotony: $Q(A,B)$ and $B \subseteq C$ implies $Q(A,C)$(All, Most);
- Left monotony: $Q(A,B)$ and $A \subseteq C$ implies $Q(C,B)$;
- Right anti-monotony: $Q(A,B)$ and $C \subseteq B$ implies $Q(A,C)$(No, Few);
- Left anti-monotony: $Q(A,B)$ and $C \subseteq A$ implies $Q(C,B)$.

Virtually all quantifiers that provide denotations for natural-language determiners are *conservative*. For instance, "Some A's are B's" is equivalent to "Some A's are A's that are B's." The one notable exception is Only, a fact on which much has been written and that has led some to question whether "only" really denotes a quantifier. Some

quantifiers such as All and Most, are right-monotonic: If "All *A*'s are *B*'s" and $A \subseteq C$, then also "All *A*'s are C's" (and similarly for "most").

We are all familiar with the all-important phenomenon of quantifier-dependence, in which quantifiers can be applied in sequence to a particular predicate as in, e.g., $\forall x \exists y Rxy$. We have not yet explained how this can be represented within the framework of the theory of generalized quantifiers, but it's intuitively clear how such an account needs to go. Given *R* and an object *a* from the domain, we need to form the collection R_a of all those objects *b* such that *Rba*; then $\forall x \exists y Rxy$ holds precisely when the set of *a*'s for which R_a falls under the quantifier $\exists$, falls under $\forall$. In other words, the combination of quantifiers $\forall\exists$ generates a dyadic quantifier that holds of a binary relation *R* precisely when for every *a* there is a *b* such that *Rab*.

The notion of *iteration* provides a generalization of this construction, by allowing us to combine quantifiers in order to obtain new ones (see Higginbotham and May 1981). Suppose Q_1 is a quantifier of type <*m*> (i.e., a quantifier taking *m*-place relations as arguments)[1] and Q_2 of type <*n*>. The iteration of Q_1 and Q_2, denoted by the juxtaposition $Q_1\,Q_2$ is a quantifier of type <*n* + *m*> defined as follows (where *R* is a relation of *n* + *m* places and the first two items are convenient abbreviations):

$$R_{a_1 \ldots a_m}(b_1, \ldots, b_n) \Longleftrightarrow R(a_1, \ldots, a_m, b_1, \ldots, b_n)$$
$$R^{Q_2}(a_1, \ldots, a_m) \Longleftrightarrow Q_2(R_{a_1 \ldots a_m})$$
$$Q_1Q_2(R) \Longleftrightarrow Q_1(R^{Q_2})$$

Notice that composition is *associative* (but *not* commutative), so we can write $Q_1\,Q_2 \ldots Q_n$ without worrying about parentheses.

Remarkably, the notion of iteration can be used also in the analysis of noun phrases, when these are interpreted as denoting Montagovian quantifiers. Recall that on such an account noun phrases such as "John" denote the collection of all subsets $X \subseteq D$ such that John is a member of X. According to the above account, the iteration of noun phrases now gives *arity-reducing* (projection) operators taking *n*-place relations to *m*-place relations, where $m \leq n$ (identifying the special case of 1-place relations with the corresponding sets of elements and the special case of 0-place relations with the truth values *true* and *false*). Then the value of

John(Mary (*kissed*))

is *true* if and only if John is a member of $kissed_{Mary}$, i.e., if and only if John kissed Mary, and *false* otherwise.

We now turn to a particularly important property of quantifiers, *permutation invariance*. Invariance under permutations was first identified by Tarski (1986) as characterizing *logical notions*, on the idea that such notions are independent of the subject matter and hence insensitive to permutations of the first-order domain. There is widespread consensus that this kind of invariance provides at least a *necessary* condition for a notion's logical character. Invariance under permutations can be applied to the modern notion of quantifier in order to single out those quantifiers that are *logical*. By a *permutation* of the domain *D* of objects we mean a function π that maps *D* onto itself in a one-to-one fashion. Such a permutation can be "lifted" to subsets of *D* by taking, for each set *A*, its point-wise image $\pi[A]$ under π, defined as $\{\pi(x): x \in A\}$. A quantifier Q is invariant

under permutations precisely when, for subsets A and B of D, $Q(A,B)$ holds iff $Q(\pi[A], \pi[B])$ holds (similarly for type <1> quantifiers). For instance, the existential quantifier $\exists$ is permutation invariant since $\pi[A]$ is nonempty whenever A is nonempty. On the contrary, the Montagovian quantifier John is not, as expected, permutation invariant, since π can map a set containing John to one that does not contain it.

Finally we mention the case of cardinality quantifiers, i.e., quantifiers that explicitly deal with cardinality constraints between the sets they relate. Historically, the first such quantifiers were the Rescher quantifier R and the Härtig quantifier I (introduced by Rescher (1962) and Härtig (1965), respectively). Their definitions are as follows:

- $\mathsf{R}(A,B) \Leftrightarrow |A| > |B|$;
- $\mathsf{I}(A,B) \Leftrightarrow |A| = |B|$.

These quantifiers have been extensively studied from a mathematical point of view (see Herre *et al.* (1991) for a survey). Härtig's quantifier is definable from Rescher's by putting $\mathsf{I}(A,B) \Leftrightarrow \neg\mathsf{R}(B,A) \wedge \neg\mathsf{R}(A,B)$ (which requires the axiom of choice) but not vice versa. While they are both obviously permutation invariant, neither quantifier is conservative, as the following both *fail*:

- $\mathsf{I}(A,B) \Leftrightarrow \mathsf{I}(A,A \cap B)$;
- $\mathsf{R}(A,B) \Leftrightarrow \mathsf{R}(A,A \cap B)$.

In a similar vein, one can introduce a quantifier F, the "Frege quantifier" of Antonelli (2010a), with the semantics

- $\mathsf{F}(A,B) \Leftrightarrow |A| \leq |B|$.

(The marginal advantage of F over R is that the Härtig quantifier can be directly represented without assuming the axiom of choice.)

One deceivingly natural way to read $\mathsf{F}(A,B)$ is "For every A there is a B" (see Boolos 1981). Such a reading is potentially misleading because it is tantalizingly close to something that one might want to express by means of alternating universal and existential quantifiers, while in fact F cannot be so interpreted (it is not even formalizable at the first order, as it interprets the Härtig quantifier).

At this juncture, it is important to bear in mind that, since all three cardinality quantifiers have type <1,1>, they are, from a semantical point of view, first order. This fact might come across as somewhat surprising, since, for instance, $\mathsf{I}(A,B)$ would seem to require the existence of a one-to-one function mapping A onto B, a notion that is usually regarded as second-order. In this respect it is important to introduce a distinction between what a statement of the form $Q(A,B)$ *asserts* and what it *expresses*. The statement $\mathsf{I}(A,B)$ does *express* the existence of such-and-such a function, but it does not *assert* it (asserting the existence of such a function is properly the purview of higher-order notions). From a purely semantical point of view, the Härtig quantifier is not any different from an uncontroversially first-order quantifier, such as Every. The statement $\mathsf{Every}(A,B)$ also *expresses* the existence of a purportedly higher-order entity, *namely*, a function f from A to B such that $f(a) = a$ for every $a \in A$, but it certainly does not *assert* the existence of such a function. The fact that both the Härtig quantifier I and the Aristotelian determiner Every (as well as the Rescher and Frege quantifiers) are first

order should not, of course, lead us to overlook the fact that they are vastly different in expressive power (see Herre *et al.* 1991 for details). Whereas Every does not outstrip the expressive capabilities of standard first-order logic, the Härtig quantifier allows us, for instance, to provide a complete characterization of the natural numbers, a feat which lies far beyond the bounds of first-order logic.

4 The Non-Standard View

The upshot of the last section is that both first-and second-order quantifiers depend for their meaning on having access to the collection of subsets of D. The difference lies in the fact that first-order quantifiers *express* the existence of particular subsets, whereas second-order quantifiers *assert* it. For instance, the sentence $\exists x\varphi(x)$ expresses the fact that the extension of $\varphi(x)$ is among the nonempty subsets of D; but in order to *assert* that the extension of $\varphi(x)$ is among the subsets of D, one needs a second-order quantifier, as in the following instance of the comprehension principle:

$$\exists P\ \forall x(Px \leftrightarrow \varphi(x)).$$

The fact remains that the specification of the meaning of both first-and second-order quantifiers is equally dependent on the specification of a second-order domain comprised of subsets of the first-order domain.

It has been known since the work of Henkin (1950) that the second-order quantifiers, in addition to the standard semantics, also admit of a *general*, or *nonstandard* semantics, in which the second-order domain is not given by the full power set $\mathcal{P}(D)$, but it is rather constituted by a collection D' of subsets—as long as such a collection satisfies certain natural closure conditions (see Enderton 2009). On such a nonstandard semantics, the second-order language is then interpretable into a two-sorted first-order language comprising (first-order) variables for objects and predicates along with a membership relation $\in$, thereby regaining many of the desirable properties of first-order logic. It went however apparently unnoticed until the work of Antonelli (2007) that first-order quantifiers *also* admit of such a nonstandard semantics.

Consider for instance the ordinary first-order quantifier $\exists$. On the *standard* interpretation, $\exists$ denotes the collection of *all* nonempty subsets of D. It is possible, however, to provide also a *nonstandard* interpretation according to which $\exists$ denotes an arbitrary collection of nonempty subsets (the collection itself may be empty, although its members—if any—may not). Similarly, in order to preserve the duality of $\exists$ and $\forall$, on the nonstandard interpretation $\forall$ denotes an arbitrary collection of subsets subject only to the proviso that D itself must be a member of that collection.[7] A similar account can be given also for the Frege quantifier F (which is also first order), where a statement $\mathsf{F}(A,B)$ holds on a nonstandard interpretation not just if $|A| \leq |B|$, but if in addition this fact is witnessed by an injective function $f\colon A \to B$ to be found in a given collection of functions also supplied by the interpretation (such a collection is also assumed to satisfy a number of specific constraints—see Antonelli 2010b for details).

One could naturally regard these alternatives not as specifying a "nonstandard" interpretation for, say, $\exists$ or F, but as providing the semantics for altogether different quantifiers with little, if any, connection to the standard $\exists$ and F. But it is perhaps conceptually more interesting to recast the situation in terms that make it clear that while the interpretation of the quantifier remains fixed, it is the underlying semantics

that is allowed to vary in nonstandard ways. In other words, *the same* quantifier $\exists$ or F is being provided either a standard or non-standard semantics. For simplicity, let us restrict our attention to monadic quantifiers, i.e., quantifiers of type $\langle 1, \ldots, 1\rangle$. Now a (general) interpretation is a pair (D, D'), where D is a nonempty first-order domain and $D' \subseteq \mathcal{P}(D)$ is the second-order domain. The interpretation is *standard* if $D' = \mathcal{P}(D)$. It is natural in certain contexts to assume that D is closed in some relevant sense, e.g., under first-order definability. On such an interpretation (D, D'), a type $\langle 1\rangle$ quantifier Q is a collection of members of D', a type $\langle 1,1\rangle$ quantifier is a relation over D', etc. Then we can take $\exists$, say, to select all and only the non-empty members of D'. On such an account, $\exists$ receives the standard interpretation on standard domains, and the general interpretation on nonstandard ones. Thus we can carry over wholesale the classification of section 3, automatically providing a nonstandard interpretation for each of the quantifiers mentioned therein. It should be evident that what this does is to provide an alternative interpretation for *the same* first-order quantifiers $\exists$, $\forall$, Every, Some, etc., rather than introducing new quantifiers altogether.

More importantly, the setup just outlined, in which an interpretation provides both a first-and a second-order domain, allows us to clarify the sense in which first-order quantifiers *express*, while second-order ones *assert*, the existence of certain subsets of the domain. In both cases it is clear that the interpretation of the quantifier depends on the specification of the domain D' of subsets of D; but while first-order quantifiers range directly over D' (they denote collections of members of D'), second-order quantifiers range over the power-set of D' (they denote collections of collections of members of D'). Notice that we thus have also a form of general Henkin semantics for the second-order quantifiers, where a second-order quantifier is now taken to denote a subset of the true power-set $\mathcal{P}(D')$ of D. Essentially, this amounts to building a standard interpretation of the second-order quantifier on top of a nonstandard interpretation for the first-order ones.[8]

5 Concluding Remark

We have thus come to a very general understanding of first-order quantifiers and the role they play in providing denotations for natural-language determiners. The great advance of the modern view was to identify quantifiers with collections of, or relations over the second-order domain. This account allowed a great many of such constructions to be properly recognized as quantifiers, from the Aristotelian determiners to cardinality quantifiers, to Montagovian individuals; it also allows us to draw a range of semantic distinctions among the quantifiers, including a specification (via permutation invariance) of the logical terms as a subclass. But this view turns out to be even further generalizable, by allowing the second-order domain to fall short of the full power-set of the first-order domain—a possibility that has been long acknowledged for second-order quantifiers, but not for first-order ones. First-order quantifiers, we have emphasized, are just as dependent on a prior notion of second-order domain as the second-order quantifiers, but in a different way.

Whereas second-order quantifiers assert the existence of subsets of the first-order domain having certain set-theoretic properties, first-order quantifiers merely express their existence. And it is this difference, we are proposing, that is at the heart of the difference in the order of quantifiers.

Notes

* We would like to thank Robert Hirsch for very helpful comments.

1 That concern was in finding a logical way of denoting numbers without engaging the paradoxes, that is, the project ultimately of *Principia Mathematica*. But Russell, driven by his epistemological conception, also sought to efface the distinction. By insisting that all names must be "logical"– they can only name that with which there is acquaintance—Russell was forced to analyze terms that are intuitively names as abbreviations for definite descriptions, and hence ultimately by the logical machinery of quantifcation (see Russell 1919).

2 We assume the interdefinability of existential generalization with respect to negation. Accordingly, we could have assumed that that the existential is the only quantifier, and defined the universal.

3 This tradition in linguistics stems from the seminal papers by Higginbotham and May (1981), Barwise and Cooper (1981) and Keenan and Stavi (1986). For the classic discussion of scope and logical form in natural language, see May (1985). For an insightful overview, see Glanzberg (2006).

4 The chief skeptic being Quine (1970).

5 In actuality, this result is a somewhat narrower, as it obtains only for quantifiers of type <1,1>; see below for discussion of quantifier types.

6 van Benthem (1983) introduces the notion of a "tree of determiners" as an intuitive way of characterizing the logical space of set-theoretic relations that demarcates the class of possible determiners. For an extended discussion, see Peters and Westerståhl (2006), pp. 160ff.

7 One could ask—although the question is somewhat tangential to the overall direction of this contribution—about the set of sentences that turn out to be valid on the nonstandard interpretation of $\exists$ and $\forall$. The somewhat surprising answer of Antonelli (2007) is that this is the same as the set of validities of positive free logic (the reason being that the non-standard semantics is equivalent to the outer-domain semantics of positive free logic).

8 This kind of general interpretation is somewhat narrower than that provided by Henkin semantics for the second-order quantifiers (thus yielding a stronger logic), for while every second-order interpretation obtained in this way is a general model in Henkin's sense, the converse is not necessarily true (although it's easy to see how to accommodate a non-standard interpretation for both the first-and second-order quantifiers).

References

G. A. Antonelli, Free quantification and logical invariance, *Rivista di estetica*, 33(1): 61–73, 2007 (special issue "Il significato eluso. Saggi in onore di Diego Marconi," M. Andronico, A. Paternoster, A. Voltolini eds).

G. A. Antonelli, The nature and purpose of numbers, *The Journal of Philosophy*, 107(4): 192–212, 2010a.

G. A. Antonelli, Numerical abstraction via the Frege quantifier, *Notre Dame Journal of Formal Logic*, 51(2): 161–79, 2010b.

J. Barwise and R. Cooper, Generalized quantifiers and natural language, *Linguistics and Philosophy*, 4(2): 159–219, 1981.

G. Boolos, For every *A* there is a *B*, *Linguistic Inquiry*, 12: 465–67, 1981.

H. Enderton, Second-order and higher-order logic, in E. Zalta, editor, *Stanford Encyclopedia of Philosophy*, 2009, URL: http://plato.stanford.edu.

M. Glanzberg, Quantifiers, in E. Lepore and B. Smith, editors, *The Oxford Handbook of Philosophy of Language*, pages 794–821, Oxford University Press, 2006.

H. Härtig, Über einen Quantifikator mit zwei Wirkungsbereichen, in L. Kalmár, editor, *Colloquium on the foundations of mathematics, mathematical machines and their applications*, pp. 31–36, Akadémiai Kiad—, Budapest, 1965.

L. Henkin, Completeness in the theory of types, *Journal of Symbolic Logic*, 15(2): 81–91, 1950.

H. Herre, M. Krynicki, A. Pinus, and J. Väänänen, The Härtig quantifier: a survey, *Journal of Symbolic Logic*, 56(4): 1153–83, 1991.

J. Higginbotham and R. May, Questions, quantifiers and crossing, *Linguistic Review*, 1(1): 41–80, 1981.

E. Keenan and J. Stavi, A semantic characterization of natural language determiners, *Linguistics and Philosophy*, 9: 253–326, 1986.

R. May, *Logical Form: Its Structure and Derivation*, MIT Press, Cambridge, MA, 1985.

R. Montague, English as a formal language, in R.H. Thomason, editor, *Formal Philosophy*, Yale University Press, 1974, originally published 1969.

A. Mostowski, On a generalization of quantifiers, *Fundamenta Mathematicæ*, 44: 12–36, 1957.
T. Parsons, The traditional square of opposition, in E. Zalta, editor, *Stanford Encyclopedia of Philosophy*, 2006, URL: http://plato.stanford.edu.
S. Peters and D. Westerståhl, *Quantifiers in Logic and Language*, Oxford University Press, 2006.
W. V. O. Quine, *Philosophy of Logic*, Prentice-Hall, 1970.
N. Rescher, Plurality quantification, *Journal of Symolic Logic*, 27: 373–47, 1962.
B. Russell, Knowledge by acquaintance and knowledge by description, in *Mysticism and Logic and other Essays*, Longmans, Green and Co., 1919.
A. Tarski, On the concept of logical consequence, in J. H. Woodger, editor, *Logic, Semantics, metamathematis. Papers from 1923 to 1938 by Alfred Tarski*, pages 409–20, Oxford University Press, 1956, Reprinted in 1983 by Hackett, J. Corcoran, ed.
A. Tarski, What are logical notions?, *History and Philosophy of Logic*, 7: 143–154, 1986, original text 1966; edited (with an introduction) by John Corcoran.
J. van Benthem, Determiners and logic, *Linguistics and Philosophy*, 6(4): 447–78, 1983.

3.5
GENERICS

Sarah-Jane Leslie

Introduction

Generics are statements such as "doctors heal people," "a tiger is striped," "the dodo is extinct," "a duck lays eggs,","kettles boil water," and "mosquitoes carry the West Nile virus." Generic statements express general claims about kinds and categories, rather than claims about particular individuals. In English, generics can be expressed using a variety of syntactic forms: *bare plurals* (e.g. "ducks lay eggs"), *indefinite singulars* (e.g. "a tiger is striped"), and *definite singulars* ("the dog is a mammal").

Quantified statements, such as "all dogs are mammals" or "most tigers are striped," can also be used to express general claims about kinds. However, quantified statements, unlike generics, can be used to answer the question *how much* or *how many*. For example, if asked "How many tigers are striped?", one could felicitously reply by saying "all/most tigers are striped." One could not felicitously reply by using the generic "tigers are striped" (Carlson 1977).

The truth conditions of generics have proved quite elusive for semanticists. For example, "dogs are mammals" seems to require for its truth that all (possible) dogs be mammals. "A tiger is striped" or "ravens are black," however, are somewhat more forgiving, since they are compatible with the existence of a few stripeless tigers (as Siegfried and Roy's performances attest), and white albino ravens. "Ducks lay eggs" and "a lion has a mane" are more forgiving still; these generics are true even though it is only the mature members of one gender which possess the relevant properties. This truth conditional laxity is limited in scope, however: We do not accept "ducks are female" or "lions are male," even though every egg laying duck is a female duck, and similarly *mutatis mutandis* for maned lions. Finally, we accept "mosquitoes carry the West Nile virus," even though fewer than 1 percent of mosquitoes carry the virus, while also rejecting "books are paperbacks," when over 80 percent of books are paperbacks. The correct analysis of the truth conditions for generics is a matter of great controversy among theorists working on the problem.

Generics vs. Existentials

The interpretation of sentences containing bare plurals, indefinite singulars, or definite singulars can be either generic as in (1) existential/specific as in (2):

(1) Tigers are striped
A tiger is striped
The tiger is striped.

(2) Tigers are on the front lawn
A tiger is on the front lawn
The tiger is on the front lawn.

The subjects in (1) are prima facie the same as in (2), yet their interpretations in (1) are intuitively quite different from those in (2). In (2) we are talking about some particular tigers, while in (1) we are saying something about tigers in general.

There are some tests that are helpful in distinguishing these two readings. For example, the existential interpretation is *upward entailing*, meaning that the statement will always remain true if we replace the subject term with a more inclusive term. For example, if it is true that tigers are on the lawn, then it will also be true that animals are on the lawn. This is not so if the sentence is interpreted generically. For example, it is true that tigers are striped, but it does not follow that animals are striped (Lawler 1973; Laca 1990; Krifka et al. 1995). Another test concerns whether we can insert an adverb of quantification (in the sense of Lewis 1975,) with minimal change of meaning (Krifka et al. 1995). For example, inserting "usually" in the sentences in (1) (for example "tigers are usually striped") produces only a small change in meaning, while inserting "usually" in (2) dramatically alters the meaning of the sentence (for example "tigers are usually on the front lawn). (For generics such as "mosquitoes carry malaria," the adverb "sometimes" is perhaps better used than "usually.")

Stage-Level and Individual-Level Predicates

We might wonder why the interpretations of (1) are so different from the interpretations of (2). The most prominent explanation is that, at least for bare plural (BP) generics, the interpretations depend on whether the predicates in question are *stage-level* or *individual-level*. In particular, stage-level predicates are thought to give rise to existential interpretations, while individual-level predicates give rise to generic ones (Carlson 1977).

Intuitively, the distinction between these two types of predicates has to do with whether the predicate denotes a property that may well be had fleetingly (making for a stage-level predicate) versus a property that is more stable and longlasting (making for an individual-level predicate). Examples of stage-level predicates include the predicates 'is drunk', 'is barking', and 'is on the lawn'—these properties are normally had only temporarily, or at least intermittently. Individual-level predicates express more stable and persistent properties, e.g. 'is tall', 'is a mammal', and 'is female'.

The distinction between stage-level and individual-level predicates can also be drawn on the basis of their differential acceptability in a variety of constructions. For example, only stage-level predicates can be used to fill in "_" in "A saw B _". That is, we can say "John saw Mary drunk/barking/on the lawn," but not "John saw Mary tall/a mammal/female." Similarly, only stage-level predicates can be used in the construction "there are Ks _": "there are firemen available/outside the window," but not "there are firemen male/tall" (Milsark 1974; Carlson 1977; Stump 1985).

The details of exactly how these different types of predicate may give rise to generic vs. nongeneric interpretations of these sentences is beyond the scope of this entry, as

the proposals tend to involve technical issues in syntax and linguistic semantics. The original proposal is due to Greg Carlson (1977), however Carlson himself later came to reject the basics of his account (Carlson 1989). More recent discussion can be found in Diesing (1992), Chierchia (1995), Kratzer (1995), Cohen and Erteschik-Shir (1997, 2002), Kiss (1998), and others.

Generics and Logical Form

There is some controversy in semantics as to the details of the logical form of generic sentences, however there is also general (though not universal) agreement about the broad features of these logical forms. This section highlights the points of general agreement (see Chapter 1.3 Logical Form).

D-Generics and I-Generics

It is helpful to separate out two categories of generic statements. Our first category includes statements such as "tigers are striped," "ravens are black," "a lion has a mane," and "the dog is carnivorous," These statements are naturally thought of as expressing generalizations about individual members of the kind. For example, we might suppose that "tigers are striped" is made true by enough individual tigers possessing stripes. The exact nature of these generalizations is highly controversial.

In our second category, we have statements like "dinosaurs are extinct," "the dodo is extinct," "tigers are widespread," and "the domestic cat is common," which are often thought to predicate a property directly of the kind in question, rather than expressing generalizations concerning its members. For example, in saying "dinosaurs are extinct," one says something about the kind *dinosaur*: namely that that kind of thing is extinct. We may notice that it is not possible to say of an individual dinosaur Dino that Dino is extinct, since only a kind can be extinct.

Examples of this second category are often referred to as D-generics ('D' for 'definite'), while examples of the first are known as I-generics ('I' for 'indefinite) (Krifka 1987). It should be noted that the indefinite singular form cannot be used to express a D-generic: Statements such as "a dinosaur is extinct" and "a tiger is widespread" are not felicitous.

Much of the work on the semantics of generics has been focused on I-generics, since they have proved the most elusive. It is widely accepted that D-generics are singular statements which predicate properties directly of kinds. For example, "tigers are extinct" predicates the property of *being extinct* directly of the kind *Panthera tigris*, and would be true just in case *Panthera tigris* had the property of being extinct (Krifka et al. 1995). The semantics of I-generics have proved much less tractable.

The Logical Form of I-Generics

For the remainder of this chapter, we will be concerned only with I-generics, since they have received the most attention, and been the subject of most controversy. Henceforth, I will use the term 'generics' to mean specifically I-generics. This section discusses the dominant theory of the logical form of I-generics, and of necessity involves some technical material. For readers unfamiliar with the general framework, it will be helpful to first consult the chapters on Logical Form (1.3), Quantifiers and Determiners (3.4), and Adverbs (3.9). It is generally agreed that that there is a two-place operator, usually

termed *Gen*, which functions as an adverb of quantification (Lewis 1975; see Chapter 3.9 on Adverbs). Examples of adverbs of quantification include 'usually', 'generally', 'sometimes', and so on. They relate one set of conditions containing at least one free variable to another set. The variables may range over individuals, or over cases or situations, as in (3), whose Lewisian analysis is given by (4):

(3) Usually, when John comes home, he sleeps
(4) Usually s [John comes home in s] [John sleeps in s]

This yields a *tripartite structure*, consisting of a *quantifier* ("Usually s"), a *restrictor* ("John comes home in s"), and a *scope* ("John sleeps in s"). In this example, the material from the when-clause is in the restrictor, and the material from the main clause is in the scope.

Generics are standardly assimilated to this model. A common assumption (Heim 1982; Kamp 1981; Kamp and Reyle 1993; Diesing 1992; Kratzer 1995; and many others) is that indefinites such as bare plurals and indefinite singulars contribute predicates with unbound variables to the logical forms containing them. During syntactic processing, the material in the sentence is divided up into the restrictor and scope (e.g. Diesing 1992). If the sentence contains a quantificational adverb (e.g. 'usually', 'always'), then any unbound variables in the restrictor are bound by that quantificational adverb, as in example (4) above. However, if there is no quantificational adverb, the generic operator Gen is introduced to bind those variables. Thus, to take a simple example, the logical form of a sentence like "ravens are black" may be given as follows:

(5) Gen x [Ravens(x)] [Black(x)]

Even theorists who do not agree with the assumption that indefinites contribute just predicates and unbound variables to their logical forms (see, for example, Chierchia 1998) tend to agree that a tripartite structure nonetheless is the correct analysis for generics. (However, Liebesman (2011) disputes this claim, arguing that I-generics should have the same logical forms as D-generics.)

A significant virtue of the tripartite structure is that it is readily able to accommodate intuitions of ambiguity, such as those associated with sentences such as "typhoons arise in this part of the Pacific," which can be understood as either "typhoons in general have a common origin in this part of the Pacific," or as "there arise typhoons in this part of the Pacific" (Carlson 1989). These different readings correspond to two different logical forms, given schematically by (6) and (7) respectively:

(6) Gen x [x is a typhoon] [x arises in this part of the Pacific]
(7) Gen e [e is a (contextually relevant) event involving this part of the Pacific] [e involves the arising of typhoons]

The term 'generic' is sometimes extended to sentences such as "Mary smokes after work," since these habitual statements share various features with generics, although they do not express generalizations concerning *kinds*. It is often believed that these statements should be analyzed with Gen (for example, Lawler 1972; Schubert and Pelletier 1989; Krifka et al. 1995; and many others):

(8) Gen e [Relevant-event-involving-Mary(e) & Occurs-after-work(e)] [Event-of-Mary-smoking(e)]

For the remainder of this chapter, the term 'habitual' rather than 'generic' will be used for these statements.

The Semantics of Generics

A great deal of work has been done on the semantics of (I-)generics, particularly on bare plural (I-)generics. It is easy to see why this is so: "Ducks lay eggs" is a true generic, while "ducks are female" is false, yet it is only the female ducks who ever lay eggs. "Mosquitoes carry the West Nile virus" is true, and "books are paperbacks" is false, yet less than one percent of mosquitoes carry the virus, while over 80 percent of books are paperbacks. How are we to account for these puzzling facts?

It is clear that generics are not equivalent to universal statements, but rather permit exceptions—that is, generics can be true even if some (or sometimes many) members of the kind lack the property in question. Generics also do not mean 'most'; it is false that most mosquitoes carry the West Nile virus and true that most books are paperbacks, but our intuitions about the truth/falsity of the corresponding generics are reversed.

As these examples suggest, Gen cannot be analyzed as sharing a meaning with any of the standard quantifiers. One question that comes up immediately is whether Gen can be considered a quantifier in any sense. Carlson (1977) argued informally that it could not, since generics do not tell us *how much* or *how many*. He notes that, if asked "how many tigers are striped," one could reply "most/all/many/some tigers are striped," but not simply "tigers are striped." Leslie (2007) offers a more formal argument to the effect that Gen cannot be considered a quantifier in any of the standard senses, since it is not permutation invariant (see Chapter 3.4 Quantifiers and Determiners), nor does it partition the power set of the domain of discourse.

A significant number of theories concerning the meaning of generics have been offered over the years. Most theories have focused primarily on bare plural generics, though some theories are intended to cover indefinite and definite singulars too. I will conclude with some remarks concerning issues specific to indefinite and definite singulars.

Possible Worlds and Normalcy-Based Approaches

Many accounts of generics are framed in terms of some or other type of quantification over possible worlds (Dahl 1975; Delgrande 1987; Schubert and Pelletier 1989; Asher and Morreau 1995; Krifka et al. 1995; Pelletier and Asher 1997; Greenberg 1998; and others; also see Chapter 2.6 Possible Worlds Semantics). Often, these possible worlds are employed to capture the intuition that generics tell us something about what is *normal* for members of a kind (see Nickel 2008). For example, it is natural to think that a generic such as "tigers are striped" tells us something about normal tigers; the only exceptions to it are those tigers who are albino, and so in some respect out-of-step with the norm for the kind. Similarly, "dogs have four legs" may strike us as true because the only dogs who do not have four legs either have birth defects or have met with misfortune.

Possible worlds are helpful here, because they allow us to consider, say, worlds in which things go *as normally as possible* for a given tiger, even if life is *actually* quite abnormal for that tiger. For example, Asher, Morreau, and Pelletier argue that "Ks ϕ" is true iff for each individual K, the most normal worlds for that K (according to a

contextually determined ordering base, see Chapter 2.6 Possible Worlds Semantics), are such that that K ϕs (Asher and Morreau 1995; Pelletier and Asher 1997). Glossed in more intuitive terms, the account states that a generic "Ks ϕ" is true iff each individual K would have the property ϕ if all was to go as normally as possible for that K. Thus, while in actuality some dogs are three-legged, one might suppose that, *had things gone more normally for each of those dogs*, they would have had four legs.

Pelletier, Asher, and Morreau connect their semantic account of generics to the literature on defeasible validity (McCarthy 1986; Reiter 1987, and others). A set of premises defeasibly or *nonmonotonically* entails a conclusion if the conclusion is likely or reasonable given the premises, yet it is nonetheless possible for the premises to be true and the conclusion false. They argue that there are many defeasibly valid patterns of entailment that involve generics, and that their semantics explains why this is so. For example, if Fido is a dog and dogs have four legs, then the inference that Fido has four legs is held to be defeasibly valid. It is *possible* that Fido is a three-legged dog, but it is still *reasonable* to suppose that Fido is four-legged on the basis of the premises. The reasonableness of this inference is explained on Pelletier, Asher, and Morreau's semantics—the generic "dogs have four legs" guarantees that, if things go normally for a given dog, then that dog will have four legs. Since we have been given no reason to suppose that Fido is abnormal, one may defeasibly conclude that Fido is four-legged.

Indisputably, a large number of true generics of the form "Ks ϕ" are such that, if things go normally for a given K, then it has the property ϕ. However, there would seem to also be a number of true generics which do not fit this profile: "ducks lay eggs," "lions have manes," "mosquitoes carry the West Nile virus," and "sharks attack swimmers." The latter two generics present a significant difficulty for normalcy-based accounts, since it is surely not *normal* for a mosquito to carry the virus, nor is it *normal* for a shark to attack a bather. For generics such as "ducks lay eggs" and "lions have manes," several theorists have argued that the domain of discourse is restricted so that we are talking about only the female ducks and the male lions respectively, and thus aim to rescue normalcy-based approaches from this criticism (for example Pelletier and Asher, 1997), since perhaps it is abnormal for a given female duck to fail to lay eggs. Leslie (2008) argues that if such domain restriction were available, then a number of false generics would be predicted to be true. Khemlani, Leslie, and Glucksberg (2009) report empirical evidence that further suggests that people do not understand these generics to involve domain restriction. If one cannot appeal to domain restriction to handle "ducks lay eggs," then such generics would appear to constitute counterexamples to this sort of approach.

Domain Restriction: Relevant Quantification and Situation Semantics

Appeals to domain restriction are not limited to normalcy-based approaches. In light of generics such as "ducks lay eggs," some theorists have argued that generics involve quantification over *relevant* individuals (actual or possible), where context determines which individuals are relevant (e.g. Schubert and Pelletier 1987; Declerk 1991; Chierchia 1995). On such views, when we consider a generic such as "ducks lay eggs," only the mature, fertile female ducks enter into our evaluation of the sentence, because, for example, they are the only potential egg-layers. The question that arises is how exactly to determine which individuals are relevant. Schubert and Pelletier (1987) offer a detailed discussion of how some of these restrictions arise—for example, via presupposition, focus, linguistic context, or explicit restriction as appropriate.

Ter Meulen (1986), Gerstner-Link (1988), and Cavedon and Glasbey (1994) offer treatments of generics in the framework of situation semantics (Barwise and Perry 1983). On such views, generics express *constraints* on situations—for example, "tigers are striped" expresses the constraint that every situation involving a tiger involves a striped tiger. Constraints are not reducible to the properties of individuals (Cavedon and Glasbey 1994). The interpretation of a generic on these views is relative to a given context (channel), so as on the above views, generics such as "ducks lay eggs" are evaluated only with respect to the female ducks (Cavedon and Glasbey 1994).

As noted above, there are philosophical and empirical difficulties (e.g. Leslie 2008; Khemlani et al. 2009; see also Cimpian, Gelman, and Brandone 2010) facing any accounts that rely on domain restriction to handle generics such as "ducks lay eggs." For these accounts to succeed, these difficulties would need to be addressed in more detail.

Stereotypes and Prototypes

A somewhat different approach to the semantics of generics is taken by theorists who suppose that generics express stereotypes or prototypes. On such views, "tigers are striped" would express that the stereotypical or prototypical tiger is striped, and likewise, "sharks attack bathers" would express a belief about the stereotypical or prototypical shark. Geurts (1985) and Declerk (1986) suggest that generics may be interpreted as expressing culturally accepted stereotypes, and Platteau (1980), Nunberg and Pan (1975), and Heyer (1985, 1990) claim that generics express prototypes, in the sense of Rosch (1978). Thus on such views, "tigers are striped" would express the fact that the stereotypical or prototypical tiger is striped.

A general concern with views of this type is that they would seem to make the holding of the relevant sort of false belief sufficient to render the generic true. For example, suppose that people falsely associate sliminess with snakes in the relevant way (for example, as a culturally held stereotype, or part of the Roschean prototype, etc.). This does not make the generic "snakes are slimy" true (Krifka et al. 1995). While it is plausible that people often *assert* generics on the basis of their prototypical/stereotypical beliefs, it is implausible to think that these beliefs enter into the *truth-conditions* of generics. Stereotypical beliefs can be, and often are, false (see Chapter 6.4.2 Language and Race).

Probabilitistic Approaches

Cohen (1996, 1999, 2004) argues that generics can be understood in terms of comparative probabilities. There are two different ways for a generic to be true on Cohen's view. The first way (Absolute) can be illustrated by "tigers are striped." This is a true generic because (roughly speaking) a randomly chosen tiger is more likely than not to be striped. The second way (Relative) a generic can be true involves comparison with other kinds. For example, on Cohen's account, "mosquitoes carry the West Nile virus" is true because (again roughly speaking) if we pick a mosquito and another insect at random, the mosquito is more likely than the other insect to carry the West Nile virus. Thus on Cohen's view, generics are made true (or false) by such probabilistic considerations. More precisely, Cohen's categories are as follows:

Absolute generics:

> 'Ks are F' is true iff the probability that an arbitrary K that satisfies some predicate in Alt(F) satisfies 'is F' is greater than .5.

Relative Generics:

> 'Ks are F' is true iff the probability that an arbitrary K that satisfies some predicate in Alt(F) satisfies 'is F' is greater than the probability that an arbitrary member of Alt(K) that satisfies some predicate in Alt(F) satisfies 'is F'.

As it stands, however, Cohen's account incorrectly predicts that generics such as "bees are sterile" will be true, since the probability that a given bee is sterile is greater than .5. To deal with such cases, Cohen introduces the homogeneity constraint:

Homogeneity Constraint:

> The above probability conditions (exceeding .5, or exceeding that of the arbitrary alternative to the kind) should hold in *all salient partitions* of the kind.

Generics such as "bees are sterile" are now predicted to be false, because there is a salient partition of bees into workers, queens, and drones—and queen bees have a very low probability of being sterile.

Leslie (2007, 2008) argues that Cohen's account faces some counterexamples. For example, humans are more likely to suffer from autism than other mammals, and so "humans are autistic" would seem to be falsely predicted to be a true (relative) generic (and it is hard to see how homogeneity could undermine this prediction). This example illustrates that unique possession of a property by a kind is not generally sufficient for the truth of a generic, yet the category of relative generic predicts that, modulo homogeneity, this should suffice. Leslie (2007, 2008) offers additional counterexamples to Cohen's conditions.

Recent empirical work also suggests that the homogeneity constraint does not guide people's understanding of generics. For example, Cimpian, Gelman, and Brandone (2010) found that adults are not at all reluctant to accept a generic that involves a property found only in one salient partition of a kind, contra the predictions of the homogeneity constraint (though this was not the intent behind the experiment).

Generics and Psychology

Recent work in psychology by Susan Gelman and her collaborators shows that generics are very easily acquired by young children. In particular, generics are understood by young children more easily than explicit quantifiers such as "most", and even "all" and "some," at least when used to make kind-wide generalizations (Gelman 2003, 2004; Gelman , Goetz, Samecka, and Flukes 2008, Gelman and Tardiff 1998; Hollander, Gelman, and Star 2002; Leslie and Gelman forthcoming; Papafragou and Schwarz 2005/2006 and other studies in progress). This is *prima facie* puzzling since, as the above discussion indicates, the semantics of generics look to be very complex.

Leslie (2008) argues from these considerations and others that generics give voice to cognitively default generalizations, and that quantified statements give voice to cognitively more sophisticated, nondefault ones. This hypothesis has gained some empirical support from the finding that both adults and young children will *interpret* and *recall*

quantified statements as generics (Hollander et al. 2002; Khemlani, Leslie, Glucksberg, and Rubio-Fernandes 2007; Leslie, Khemlani, and Glucksberg 2011; Leslie and Gelman forthcoming).

Leslie further argues that once we understand the role that generics play in our psychology, we can develop an account of when generics are true and false. Leslie divides generics into three categories: characteristic (including items such as "ducks lay eggs"), majority (for example "cars have radios") and striking/dangerous (for example "mosquitoes carry the West Nile virus"). These different classes have different requirements that the world must meet for the corresponding generic to be true. Cimpian, Brandone, and Gelmen (2010) report empirical results that support Leslie's thesis that these particular factors differentially impact on adults' willingness to accept generic statements (for example adults accept generics at lower prevalence levels if the property in question is strikingly dangerous).

Definite and Indefinite Singular Generics

The above discussion was primarily centered on accounts of bare plural generics, which have received the most discussion in the literature. Singular generics introduce their own sets of complications. Unlike bare plurals, singular generics can easily be infelicitous. For example, it is perfectly fine to say (9) or (10):

(9) A madrigal is polyphonic
(10) A football hero is popular

But not (11):

(11) *A madrigal is popular

(Notice, however, that the plural version of (11) "madrigals are popular" is perfectly felicitous (though perhaps, unfortunately, false).) Lawler (1973) notes that indefinite singulars are only felicitous when they express properties that are somehow "necessary", "essential" or "inherent" to the kind. Burton-Roberts (1977) argues that indefinite singulars carry a special normative force, while Krifka et al. (1995) take a different tact, arguing that the facts can be accounted for in terms of logical form. Greenberg (1998) and Cohen (2001) argue that indefinite singulars can express only "rules and regulations," in the sense of Carlson (1995). These analyses are often implemented by way of possible worlds semantics.

Definite singular generics appear to invoke constraints similar to those of indefinite singulars, but also some more besides. For example, definite singular subjects are often infelicitous if they do not refer to well-established kinds (Krifka et al. 1995; example from Carlson 1977, attributed to Barbara Partee):

(12) The coke bottle has a narrow neck
(13) *The green bottle has a narrow neck

Other constraints on the acceptability of definite singulars have been noted by Nunberg and Pan (1975), Carlson (1977), Dahl (1985), and others. However, definite singular generics have received less attention overall than indefinite singulars, while bare plurals have received by far the most discussion in the literature.

New Directions

Recently, several philosophers have begun to consider the role generics, and the generalizations they express, may play in various forms of social prejudice (for example Haslanger forthcoming; Leslie forthcoming; see Chapter 6.4.2 Language and Race for more details). More generally, generics continue to be a topic of considerable interest to linguists, philosophers, and more recently, psychologists. There has recently been an explosion of interest among psychologists concerning generics (e.g. Gelman 2003; Prasada and Dillingham 2006, 2009). Recent and ongoing work in psychology has been examining issues such as how children acquire generics, how children and adults process generics, the role of generics in reasoning, and the influence of generics in various forms of prejudice. This new range of empirical findings will surely be a significant influence on the topic going forward.

Related Topics

2.6 Possible-Worlds Semantics
3.4 Quantifiers and Determiners
3.8 Plurals
3.9 Adverbs
3.10 Mass Terms
6.3.1 The Language of Causation
6.4.1 Language, Gender, and Sexuality
6.4.2 Language and Race.

References

Asher, N. and M. Morreau. (1995) "What some generic sentences mean". In G. Carlson and F. J. Pelletier (eds.) *The Generic Book*. Chicago: Chicago University Press, pp. 300–339.

Barwise, J., and Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Burton-Roberts, N. (1977). "Generic sentences and analyticity." *Studies in Language*, 1, 155–196.

Carlson, G. (1977). Reference to Kinds in English. Ph.D. dissertation, University of Massachusetts, Amherst.

Carlson, G. (1989). 'On the semantic composition of English generic sentences'. In G. Chierchia et al. (eds.), *Properties, Types and Meanings, Vol. II: Semantic Issues*, Dordrecht: Kluwer, pp. 167–192.

Carlson, G. (1995). "Truth-conditions of generic sentences: Two contrasting views." In G. Carlson and F. J. Pelletier (eds.) *The Generic Book*. Chicago: Chicago University Press, pp. 224–237.

Cavedon, L., and Glasbey, S. (1994). "Outline of an information-flow model of generics." *Acta Linguistica Hungarica*, 42.

Chierchia, G. (1995). "Individual level predicates as inherent generics". In G. Carlson and F. J. Pelletier (eds.) *The Generic Book*. Chicago: Chicago University Press, pp. 176–224.

Chierchia, G. (1998). "Reference to kinds across languages." *Natural Language Semantics*, vol. 6.

Cimpian, A., Brandone, A. C., and Gelman, S. A. (2010). "Generic statements require little evidence for acceptance but have powerful implications." *Cognitive Science*, 34(8), 1452–1482.

Cimpian, A., Gelman, S. A., and Brandone, A. C. (2010). "Theory-based considerations influence the interpretation of generic sentences." *Language and Cognitive Processes*, *25*(2), 261–276.

Cohen, A. (1996). Think Generic: The Meaning and Use of Generic Sentences. Ph.D. dissertation, Carnegie Mellon University.

Cohen, A. (1999). "Generics, frequency adverbs and probability." *Linguistics and Philosophy*, 22, 221–253.

Cohen, A. (2001) "On the generic use of indefinite singulars." *Journal of Semantics*, 18(3), 183–209.

Cohen, A. (2004). "Generics and mental representation." *Linguistics and Philosophy*, 27(5), 529–556.

Cohen, A., and Erteschik-Shir, N. (1997). "Topic, focus and the interpretation of bare plurals." In P. Dekker, M. Stokhof, and Y. Venema (eds.) *Proceedings of the 11th Amsterdam Colloquium*, pp. 31–36.

Cohen, A., and Erteschik-Shir, N. (2002). "Topic, focus, and the interpretation of bare plurals." *Natural Language Semantics*, 10(2), 125–165.
Dahl, O. (1975). "On generics." In E. L. Keenan (ed,) *Formal Semantics of Natural Language*,. Cambridge: Cambridge University Press, pp. 99–111.
Dahl, O. (1985). *Tense and Aspect Systems*. Oxford: Blackwell.
Declerk, R. (1986). "The manifold interpretations of generic sentences." *Lingua*, 68, 149–188.
Declerk, R. (1991). "The origins of genericity." *Linguistics*, 29, 79–101.
Delgrande, J. P. (1987). "A first-order conditional logic for prototypical properties." *Artificial Intelligence*, 33, 105–130.
Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.
Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. New York: Oxford University Press.
Gelman, S. A. (2004). "Learning words for kinds: Generic noun phrases in acquisition." In D. G. Hall and S. R. Waxman (eds.) *Weaving a Lexicon*. Cambridge, MA: MIT Press.
Gelman, S. A., Goetz, P. J., Sarnecka, B. S., and Flukes, J. (2008). "Generic language in parent-child conversations." *Language Learning and Development*, 4, 1–31.
Gelman, S. A., & Tardif, T. Z. (1998). "A cross-linguistic comparison of generic noun phrases in English and Mandarin." *Cognition*, 66, 215–248.
Gerstner-Link, C. (1988). Uber Generizitat. Generische Nominalphrasen in Singularen Aussagen und Generischen Aussagen. Ph.D. dissertation, University of Munich.
Geurts, B. (1985). "Generics." *Journal of Semantics*, 4.
Greenberg, Y. (1998), *Manifestations of Genericity*. Routledge: New York.
Haslanger, S. (Forthcoming), "Ideology, generics, and common ground", in Charlotte Witt (ed.) *Feminism and Metaphysics*.
Heim, I. (1982). The Semantics of Definite and Indefinite Noun Phrases. Ph.D. dissertation, University of Massachusetts, Amherst.
Heyer, G. (1985). "Generic descriptions, default reasoning, and typicality." *Thoretical Linguistics*, 12, 33–72.
Heyer, G. (1990). "Semantics and knowledge representation in the analysis of generic descriptions." *Journal of Semantics*, 7, 93–110.
Hollander, M. A., Gelman, S. A., and Star, J. (2002). "Children's interpretation of generic noun phrases." *Developmental Psychology*, 38, 883–894.
Kamp, H. (1981), "A theory of truth and semantic interpretation". In J. Groenendijk, J. Janssen and M. Stokhof (eds.) *Formal Methods in the Study of Language*. Forts Publications. Dordrecht, pp. 277–322.
Kamp, H. and U. Reyle. (1993). *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.
Khemlani, S., Leslie, S. J., Glucksberg, S., & Rubio Fernandez, P. (2007). "Do Ducks Lay Eggs? How Humans Interpret Generic Assertions." *Proceedings of the 29th Annual Cognitive Science Society*. Nashville, TN: Cognitive Science Society.
Khemlani, S., Leslie, S. J., and Glucksberg, S. (2009). "Generics, prevalence, and default inferences". *Proceedings of the 31st Annual Cognitive Science Society*. (2009) Amsterdam: Cognitive Science Society.
Kiss, K. E. (1998). "On generic and existential bare plurals and the classification of predicates." In S. Rothstein, (ed.) 1998. *Events and Grammar*. Dordrecht: Kluwer, pp. 145–162.
Kratzer, A. (1995). "Stage-level and individual-level predicates". In G. Carlson and F. J. Pelletier (eds.) *The Generic Book*. Chicago: Chicago University Press, pp. 125–176.
Krifka, M. (1987). *An Outline of Genericity*, partly in collaboration with Claudia Gerstner. SNS-Bericht 87-23, University of Tubingen.
Krifka, M., F. Pelletier, G. Carlson, A. Ter Meulen, G. Chierchia and G. Link. (1995). "Genericity: An introduction". In G. Carlson and F. J. Pelletier (eds.) *The Generic Book*. Chicago: Chicago University Press, pp. 1–125.
Laca, B. (1990). "Generics objects: Some more pieces of the puzzle." *Lingua*, 81, 25–46.
Lawler, J. (1972). "Generic to a fault." *CLS* 8, Chicago: Chicago Linguistic Society, 247–258.
Lawler, J. (1973). "Studies in English generics." *University of Michigan Papers in Linguistics* 1(1).
Leslie, S. J. (2007). "Generics and the structure of the mind." *Philosophical Perspectives*, 21(1), 375–403.
Leslie, S. J. (2008). "Generics: Cognition and acquisition." *Philosophical Review*, 117(1), 1-47.
Leslie, S. J. (forthcoming). "The original sin of cognition: Fear, prejudice, and generalization." *The Journal of Philosophy*.
Leslie, S. J., and Gelman, S. A. (forthcoming). "Quantified statements are recalled as generics: Evidence from preschool children and adults". *Cognitive Psychology*.

Leslie, S. J., Khemlani, S., and Glucksberg, S. (2011). "All ducks lay eggs: The generic overgeneralization effect". *Journal of Memory and Language*, 65, 15–31.

Lewis, D. (1975). "Adverbs of quantification". In E. L. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press, pp. 3–15.

Liebesman, D. (2011), "Simple generics", *Noûs*, 45(3), 409–442.

McCarthy, J. (1986). "Applications of circumscription to formalizing commonsense knowledge." *Artificial Intelligence*, 28, 89–116.

ter Meulen, A. (1986). "Generic information, conditional contexts and constraints." In *On Conditionals*, E. C. Traugott, A. ter Meulen, J. S. Reilly, and C. A. Ferguson (eds.) Cambridge: Cambridge University Press.

Milsark, G. (1974). Existential Sentences in English. Ph.D. dissertation, MIT.

Nickel, B. (2008), "Generics and the way of normality." *Linguistics and Philosophy*, 31 (6), 629–648.

Nunberg, G. and Pan, C. (1975). "Inferring quantification in generic sentences." In *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society*.

Papafragou, A. and Schwarz, N. (2005/6). "*Most* wanted." *Language Acquisition* 13 (Special Issue: On the Acquisition of Quantification) 207–251.

Pelletier, F. and Asher, N. (1997). "Generics and defaults". In J. van Benthem and A. ter Meulen (eds.) *Handbook of Logic and Language*. Cambridge, MA: MIT Press, pp. 1125–1179.

Platteau, F. (1980). "Definite and indefinite generics." In J. van der Auwera (ed.) *The Semantics of Determiners*. London: Croom Helm, pp. 112–123

Prasada, S. and Dillingham, E.M. (2006). "Principled and statistical connections in common sense conception." *Cognition*, 99, 73–112.

Prasada, S. and Dillingham, E.M. (2009). "Representation of principled connections: A window onto the formal aspect of common sense conception." *Cognitive Science* 33, (3), 401–448.

Reiter, R. (1987). "Nonmonotonic reasoning." In *Annual Review of Computer Science*, J. F. Traub, N. J. Nilsson and B. J. Grosz (eds.) Palo Alto: Annual Reviews Inc, pp. 147–186.

Rosch, E. H. (1978). "Principles of categorization". In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum Associates, pp. 27–48.

Schubert, L.K. and Pelletier, F.J. (1987). "Problems in representing the logical form of generics, bare plurals, and mass terms". In E. Lepore (ed.), *New Directions in Semantics*. London: Academic Press, pp. 387–453.

Schubert, L. K. and Pelletier, F. J. (1989). "Generically speaking, or using discourse representation theory to interpret generics." In G. Chierchia, , B.H. Partee, and R. Turner, (eds.) *Properties, Types and Meaning*. Dordrecht: Kluwer, pp. 193–268.

Stump, G. (1985). *The Semantic Variability of Absolute Constructions*. Dordrecht: Reidel.

3.6
ANAPHORA

Jeffrey King

To say *precisely* what anaphora is in a way that is theory neutral is nontrivial. Very roughly, cases of anaphora involve one expression, the *anaphoric* one, being interpreted in light of another expression, its *antecedent*. A paradigmatic case would be the following:

1. Joanna$_1$ came to talk to me. She$_1$ was upset.
   (I'll use numerical coindexing to indicate intended anaphoric relations.)

Here 'She' is the anaphor (*anaphoric on* 'Joanna') and 'Joanna' is its antecedent. It is widely held that expressions of many different syntactic categories produce anaphora. However, for purposes of the present chapter, we shall concentrate on pronominal anaphora since it is the most widely studied form of anaphora and has generated the most controversy.

The reason for interest in anaphoric pronouns is that it is sometimes not clear how they are functioning semantically. Before turning to such cases, it is worth saying a few words about cases in which it *is* relatively clear how anaphoric pronouns are functioning semantically. It is generally agreed that in sentences like 1, the pronoun simply picks up its antecedent's semantic value. For example, those who think that the name 'Joanna' is a referring expression whose semantic value is Joanna will hold that the semantic value of the anaphoric pronoun is Joanna as well. Hence the pronoun is simply a referring expression. No one has lost any sleep attempting to understand cases of anaphora illustrated by examples like 1. Similarly, there is wide agreement that in certain cases, anaphoric pronouns are functioning semantically as bound variables do in first order logic. In cases like:

2. [Every politician]$_1$ believes she$_1$ should have more power.

it would appear that 'she' is simply a variable bound by 'Every politician', so that 2 could be represented as:

2a. [Every politician: x] believes that x should have more power.

Again here, it would seem that the functioning of the anaphoric expression is well understood.

However, as indicated above, there are other cases in which it is unclear how anaphoric pronouns are functioning semantically and for this reason such cases have

attracted lots of attention from linguists and philosophers. One such case is one in which a pronoun is anaphoric on an indefinite description in another sentence:

3. [A man]$_1$ showed up at Sue's office yesterday. He$_1$ was with the IRS.

Many philosophers and linguists think that indefinite descriptions are existential quantifiers. Further, it is generally thought that the scope of a quantifier cannot extend beyond the sentence in which it occurs. That this is so is confirmed by trying to bind pronouns in other sentences using other quantifiers:

4. *[Every/No student]$_1$ took the final. He$_1$ didn't do well.

Further, as noted by Evans (1977), if we consider examples like 3 but using different quantifiers, we see that construing the pronoun as a variable bound by its quantifier antecedent would yield the wrong truth conditions. It would make

5a. [Exactly one man]$_1$ drank champagne. He$_1$ was ill.

equivalent to

5b. Exactly one man drank champagne and was ill.

But of course this is incorrect. So in 3, we appear to have a pronoun anaphoric on a quantifier that cannot be construed as a variable bound by the quantifier. It also does not seem as though the pronoun in 3 is a referring expression. For one thing, holding that the pronoun is a referring expression would assign the discourse the wrong modal profile. Suppose exactly one man, Alan, did appear at Sue's office yesterday and that the pronoun in the second sentence refers to him. Then, consider a world in which a different man showed up at Sue's office yesterday and in which Alan works for the IRS. Both sentences of 3 would be true at such a world if the pronoun in the second sentence refers to Alan (and 'A man' in the first sentence is an existential quantifier). But this is incorrect. Further, in more complex cases it is much harder to see how pronouns anaphoric on indefinites whose scope they do not occur in can be referring expressions:

6. John believes [a man]$_1$ broke into Sue's apartment. Tom believes he$_1$ got in through the window.

Both sentences in 6 can be true even if no man did break into Sue's apartment. But then it seems as though the pronoun here cannot be a referring expression. And surely it would be desirable to have a theory on which the pronouns in 3 and 6 work in the same way semantically. So it appears that the pronoun in 3 is not a referring expression, nor is it a variable bound by its quantifier antecedent. But then how is it functioning semantically? Because the pronoun in 3 has an antecedent prior to it in the discourse, examples like this are often called instances of *discourse anaphora*. So, there is a theoretical question as to how anaphoric pronouns function semantically in cases of discourse anaphora.

A second sort of case in which it is unclear how an anaphoric pronoun is functioning semantically is a case in which *in a single sentence* a pronoun appears to be anaphoric on

a quantifier but not a variable bound by it. The primary examples of interest here are so-called *donkey sentences*. Examples include:

7a. Every man who owns [a donkey]$_1$ beats it$_1$.
7b. If [a man]$_1$ owns [a donkey]$_2$, he$_1$ beats it$_2$.

It is generally agreed that both sentences are true just in case every male donkey owner beats every donkey he owns. For obvious reasons, sentences like 7a are sometimes called *relative clause donkey sentences* and those like 7b are called *conditional donkey sentences*. The term *donkey anaphora* generally applies to both kinds of example. Given the truth conditions of these sentences, it is clear that the anaphoric pronouns here cannot be referring expressions. Further, there is reason to think they cannot be variables bound by their quantifier antecedents either. First, if we try using other quantifiers in 7a/7b we cannot bind the pronouns:

7'a. *Every man who owns [every/no donkey]$_1$ beats it$_1$.
7'b. *If [a man]$_1$ owns [every/no donkey]$_2$, he$_1$ beats it$_2$.

This is thought to be due to the fact that the scope of a quantifier in a relative clause is confined to the relative clause; and similarly for the scope of a quantifier in the antecedent of a conditional. Further, in 7b, even if the existential quantifier could take scope over the entire conditional, it doesn't seem like the resulting truth conditions could require every male donkey owner to beat *every* donkey he owns. So here again, we find anaphoric pronouns with quantifier antecedents that cannot be understood as referring or as being bound by their quantifier antecedents. Again, this raises the theoretical questions of how donkey pronouns are functioning semantically.

I now turn to the main theoretical approaches that give accounts of the semantic functioning of pronouns in discourse anaphora and donkey anaphora. Due to limitations of space, I shall be content to informally sketch the main ideas of each approach.

The first sort of approach to donkey and discourse anaphora I'll discuss, which I'll call the *descriptive approach*, claims that the anaphoric pronouns in question function semantically as definite descriptions. Descriptive approaches have been defended by Cooper (1979), Davies (1981), Neale (1990), Heim (1990), and Elbourne (2005) among others. In cases of discourse anaphora, it is easy to see why one might think that pronouns function semantically as definite descriptions. The second sentences of the following two discourses seem to mean the same thing:

8a. [A female student]$_1$ took my class and flunked it. She$_1$ was upset.
8b. [A female student]$_1$ took my class and flunked it. [The female student who took my class and flunked it]$_1$ was upset.

(8b may sound a bit awkward, but that seems to be because the descriptive material in the description isn't needed and so is redundant. When such material is needed to figure out who is being talked about, the awkwardness disappears: 'A female student took my class and flunked it but another female student took my class and did well. The female student who took my class and flunked it was upset.') Descriptive *theories*, particular versions of the descriptive approach, can differ along the following dimensions. First, it might be claimed that at the relevant level of syntax, the anaphoric pronouns in ques-

tion *are* definite descriptions (Cooper 1979, Heim 1990, Elbourne 2005) or it might be claimed that the pronouns are merely *semantically interpreted* as definite descriptions (Neale 1990). Second, such views can differ as to how the descriptive content of the description that provides the interpretation of the pronoun is generated. For example, one could hold that it is generated by being salient in context or one might claim that it is recovered in some rule-governed way from the linguistic environment of its antecedent. Finally, descriptive theories can differ in terms of the semantics they assign to definite descriptions and hence to the anaphoric pronouns to which they are semantically equivalent. For example, Neale (1990) endorses a Russellian treatment of definite descriptions on which they are quantifiers and on which the truth conditions of sentences of the form 'The F is G' require there to be exactly one F, which is also G. On this treatment such sentences entail the existence and uniqueness of an F. By contrast, Heim (1990) and, following her, Elbourne (2005) take such sentences to *presuppose* that there is a unique F and to assert that it is G. For Heim and Elbourne, if this presupposition fails we have no truth value for the sentence, whereas for Neale the failure for there to be a unique F results in falsity. For Heim and Elbourne, these anaphoric pronouns have as their semantic values functions from situations (or situations and individuals—see example 11) to individuals.

While it is easy to see in outline how descriptive theories will treat instances of discourse anaphora such as 8a, it should be not at all obvious how they will treat instances of donkey anaphora such as 7a and 7b above. For Neale (1990) and Heim (1990), 7a would essentially get regimented as follows:

7aR. ([Every x: man x & [a y: donkey y & x owns y]][the *z*: donkey *z* and x owns *z*] (x beats *z*)).

The underlined definite description corresponds to the pronoun 'it' in 7a. Note that it contains a variable bound by the quantifier 'Every man who owns a donkey.' But the problem here is that 7aR either entails (Neale 1990) or presupposes (Heim 1990) that every man who owns a donkey owns exactly one donkey. And yet native speakers tend to judge 7as true in a case in which many male donkey owners own more than one donkey, so long as all male donkey owners beat all donkeys they own. Similar problems arise with 7b.

I know of two attempts to deal with this problem in the literature. Neale [1990], following Parsons (1978) and Davies (1981), claims that there are "numberless descriptions," which he represents as 'whe F' (instead of 'the F'), such that 'Whe F is G' is true iff every F is G (and there is an F). Hence so-called numberless descriptions are essentially universal quantifiers. Interpreting the pronoun in 7a as a numberless description, we would get an LF as follows:

7aR1. ([Every x: man x & [a y: donkey y & x owns y]][whe *z*: donkey *z* and x owns *z*] (x beats *z*))

This is true iff every man who owns a donkey beats every donkey he owns. Though Neale gets the right truth conditions, positing these numberless descriptions seems a bit ad hoc. Further, if singular pronouns anaphoric on singular indefinites can have numberless or Russellian readings, there needs to be an explanation of why many such pronouns do not have numberless readings including the following:

9. A student came to see me today. She was upset.

There is no reading of the second sentence on which it entails that every student who came to see me today was upset. Neale suggests that in a case like this, one likely has a student in mind who provides the grounds for the one's utterance and so the pronoun is interpreted as a Russellian description. But his explanation doesn't work. Suppose I am telling you that there are some things I know on general grounds. You ask for an example. I say:

10. Well, a man is eating dinner in a restaurant right now in New York. He is drinking an expensive glass of wine. He has just finished a long day of work on Wall Street.

Surely, though I have no particular man in mind, what I've said is true iff there is a man having dinner at a restaurant in New York right now who is drinking a glass of expensive wine and has just finished a long day of work on Wall Street. Just as surely for example, the second sentence does not have a reading on which its entails that *every* man eating dinner right now in a restaurant in New York is drinking a glass of expensive wine. But if the pronoun *qua* description in the second sentence were given a numberless reading, this would be entailed. Here I have no man in mind in making my utterance, so by Neale's own lights the numberless readings should be available.

Heim (1990) suggested dealing with the uniqueness presupposition, generated by treating the pronouns in 7b as definite descriptions, by taking conditionals to quantify over situations. Very roughly, she claims 7b is true iff every minimal situation s in which the antecedent is true (which is a situation containing one man who owns one donkey), can be extended to a situation s' in which the consequent is true. However, when the consequent is evaluated in s', the pronouns *qua* definite descriptions are evaluated *in s*. But there is a unique man owning a donkey and a unique donkey owned by a man in s. So we just make sure that the man who owns a donkey (in s) beats *in s*' the donkey owned by a man (in s). The uniqueness presuppositions of the pronouns *qua* definite descriptions are rendered harmless due to the minimality of s.

However Heim's treatment of 7b leads to what has been dubbed *the proportion problem*. Specifically, Heim's account predicts that the following sentence is true if there is one man who owns 100 donkeys and beats them all, while 99 other men own one donkey each and don't beat them:

11. If a man owns a donkey, he usually beats it.

The reason Heim's theory predicts this is that *most* ("usually") minimal situations consisting of a single man owning a single donkey are extendable to a situation in which the man beats the donkey (the one donkey beater is "part" of 100 such situations). However, the prediction that the sentence is true in this circumstance seems incorrect. Though Heim (1990) considers ways to deal with the problem, none ultimately succeeds; nor is the matter dealt with satisfactorily in Elbourne (2005). Further, Elbourne (2005) tries to extend Heim's [1990] treatment of 7b, as quantifying over minimal situations in an attempt to render uniqueness presuppositions of the pronouns harmless, to 7a. Elbourne construes quantification generally ('Every man who owns a donkey') as quantification over minimal situations. Roughly, when we evaluate 'it' in 7a we are doing so with

respect to a minimal situation in which a man owns a donkey. But such a situation will again contain a unique donkey, rendering the uniqueness presupposition of the pronoun *qua* definite description harmless. Unfortunately, Elbourne thereby runs into a version of the proportion problem. His theory predicts that the following sentence is true if, again, there is one man who owns 100 donkeys and beats them all while 99 other men own one donkey each and don't beat them:

12. Most men who own a donkey beat it.

Again, this prediction seems clearly incorrect.

There is a second approach to discourse and donkey anaphora that is in certain ways similar to certain versions of the descriptive approach (for example that of Davies 1981 and Neale 1990), though it avoids some of its difficulties mentioned above. On these versions, the anaphoric pronouns in questions are semantically functioning as definite descriptions, which are in turn understood as quantifiers. On the Context Dependent Quantifier (CDQ) approach to discourse and donkey anaphora, suggested by Wilson [1984] and elaborated and defended in King (1987, 1994, 2004), pronouns in discourse anaphora are understood as quantifiers as well. However, they are quantifiers whose *force* (universal, existential, etc.), *restriction* (what they quantify over), and *scope* (relative to other scoped elements) are determined by features of their linguistic environments. In the case of discourse anaphora:

9. A student came to see me today. She was upset.

CDQ claims that the pronoun is a quantifier that has existential force (inherited from its antecedent), and is restricted to *students who came to see me today*. Hence the second sentence is true iff a student who came to see me today was upset. This seems intuitively correct.

As for donkey anaphora, King (2004) proposes an account of conditional donkey anaphora on which, as on Heim's (1990) account, conditionals quantify over situations. The rough idea is that in a case like

13. If Sarah owns a donkey, she beats it.

though the pronoun 'it' is an existential quantifier ranging over donkeys owned by Sarah, and hence the consequent is equivalent to Sarah beats a donkey she owns, the whole conditional is true iff every minimal situation s_1 in which Sarah owns a donkey (i.e. in which the antecedent is true) is part of situation s_2 in which Sarah beats a donkey she owns *in* s_1 (i.e. the consequent is true in s_2 in virtue of the donkey in s_1, which of course must be in s_2 too). Because King's (2004) account of donkey conditionals is so similar to Heim's (1990), he too faces the proportion problem here. See King (2004) for an inconclusive discussion of the relevant issues and attempted solutions. As to the CDQ account of relative clause donkey sentences, it straightforwardly predicts one of the readings such sentences have. See King (2004) for details.

Let me now turn to the third kind of approach to discourse and donkey anaphora: *discourse representation (DR) approaches*. Such approaches were pioneered by Heim (1982) and Kamp (1981). My exposition here will follow Kamp's implementation of the basic ideas. In processing sentences of a natural language discourse, speakers produce

representations of the discourse that Kamp calls *discourse representation structures* (DRS). As the discourse proceeds and a new sentence is uttered, it is interpreted in the context of the DRS the hearer has produced to that point. The DRS is "updated" in the processing of the new sentence yielding a new DRS. As indicated above, it is not possible here to lay out Kamp's formal theory, but we can sketch the outlines of the theory and how it handles discourse and donkey anaphora. The formal theory is comprised by the following things. First, there is a formal language in which a discourse, understood as a finite string of sentences of the language, can be formulated. Second, there are rules of DRS construction. Given the DRS at a given point corresponding to the processing of the first n sentences in the discourse, these rules tell you how to update that DRS depending on what type of sentence the n+1st sentence of the discourse is. Finally, there is a definition of what it is for a discourse D to be true in a model M relative to a complete DRS K for D.

Which construction rule applies at a given point in processing a sentence depends on which expression in the sentence "has widest scope" (i.e. was added last in constructing the sentence) among the expressions in the sentence that have not already been treated. In order to explain how DRS approaches handle discourse and donkey anaphora, we need to informally explain the construction rules for indefinite descriptions, pronouns, conditionals, and universal quantifiers. We then need to explain how the semantics applies to the DRS resulting from the application of these construction rules.

Here are some notions from Kamp's formal theory. A *discourse representation* (DR) is a pair <U,Con> , where U is a set of *discourse referents* (to be discussed below) and Con is a set of *conditions*, which are simply sentences of a language that extends the language that is used to formulate a discourse. A discourse representation structure (DRS) is a set of DRs. The simplest rules of DRS construction tell us what to add to U (if anything) or to Con in applying the rules to sentences in the DR in question. The more complex rules tell us how to add DR s to our DRS. Examples should make this clear.

Let's begin by considering an example of discourse anaphora.

14. Pedro owns [a donkey]$_1$. It$_1$ thrives.

The beginning DRS for this discourse is a set containing one DR: {<Ø, {Pedro owns a donkey, It thrives} }. Assuming that 'Pedro' has widest scope in the first sentence, the rule for proper names tells us to add a discourse referent, say 'u', to U (=Ø) and add the conditions 'u =Pedro' and 'u owns a donkey' to Con (={Pedro owns a donkey, It thrives}), yielding the following DRS: {<{u},{Pedro owns a donkey, It thrives, u=Pedro, u owns a donkey}>}. Now assume that 'a donkey' in 'u owns a donkey' has widest scope of the expressions not yet treated in the sentence. Then we apply the indefinite rule, which tells us to add another discourse referent, say v, to U(={u}) and to add the conditions 'donkey (v)' and 'u owns v' to Con (= {Pedro owns a donkey, It thrives, u=Pedro, u owns a donkey}) yielding the DRS {<{u,v}, {Pedro owns a donkey, It thrives, u=Pedro, u owns a donkey, donkey (v), u owns v}>}. At this point, nothing further can be done to the first sentence. In the second sentence, the pronoun in subject position has widest scope, and so we apply the pronoun rule to it. Roughly, this rule tells us to find a "suitable" *accessible* discourse referent, in this case that is going to be 'v', and add to Con 'v thrives', yielding the following *complete* DRS for our original discourse: {<{u,v}, {Pedro owns a donkey, It thrives, u=Pedro, u owns a donkey, donkey (v), u owns v, v thrives}>}. Now put very roughly, Kamp's (1981) semantics tells us that this DRS is true relative to model M iff

there is a function f from U (= {u,v}) into M such that f(u) = referent of 'Pedro' in M, <f(u), f(v)> ε ext_M('owns') and f(v) ε ext_M('donkey') and ext_M('thrives') (ext_M ('owns') is, of course, the extension of 'owns' in M and similarly for other predicates). In other words, the discourse is true iff Pedro owns a donkey that thrives. This seems intuitively correct.

Turning now to donkey anaphora, because Heim's and Kamp's original theories treated relative clause and conditional donkey sentences in virtually the same way, we need only consider one of these. Let's try 'Every man who owns a donkey beats it.' Here I'll be a bit less explicit about the construction. In the previous examples, all DRS were sets containing one DR. Here that will not be the case. The beginning DRS is: {<Ø, {Every man who owns a donkey beats it}>}. The 'Every man' quantifier has widest scope here, so we apply the universal quantifier rule first. This rule tells us to add to the DRS *two* new DRs, in effect corresponding to the *restriction of the quantifier* ('man who owns a donkey') and the quantifier's *nuclear scope* ('beats it'). These DRs are respectively <{u}, {man(u), u owns a donkey}> and <Ø, {u beats it}>. The entire resulting DRS is the following set of *three* DR's: {<Ø, {Every man who owns a donkey beats it}>, <{u}, {man(u), u owns a donkey}>, <Ø, {u beats it}>}. After applying the indefinite rule to the condition 'u owns a donkey' in the second DR in this DRS and applying the pronoun rule to 'it' in the third DR, the resulting DRS is: {<Ø, {Every man who owns a donkey beats it}>, <{u,v}, {man(u), u owns a donkey, donkey(v), u owns v }>, <Ø, {u beats it, u beats v}>}. Roughly, this DRS is true in M iff every function f from {u,v} into M such that f(u) ε ext_M ('man'), f(v) ε ext_M('donkey') and <f(u),f(v)> ε ext_M('owns') can be extended to a function f' such that <f'(u), f'(v)> ε ext_M('beats'). (In the current case f' won't extend f, because f' need only be defined on u and v, which f is. But in the general case f' will have to extend f in some cases.)

Thus, the semantics predicts that the original sentence is true iff every man who owns a donkey beats every donkey he owns. As I indicated above, the conditional 'If a man owns a donkey, he beats it' is treated in virtually the same manner with the conditional understood as universally quantifying over men and donkeys they own in virtue of the presence of the indefinites 'a man' and 'a donkey' in the antecedent of the conditional.

Before turning to other matters, a couple of remarks are in order. First, note how indefinites are treated on the present proposal. Thinking of discourse referents for a moment as free variables, an indefinite contributes to a DRS a free variable while putting a condition on it (donkey(u)). In our example of discourse anaphora, the indefinite has existential force only because the semantics includes default existential generalization of free variables ("a DRS is true relative to M iff *there is* a function that maps the discourse referents to elements of M . . ."). By contrast, in the relative clause donkey sentence the discourse referent/free variable introduced by the indefinite gets bound by the universal quantifier just as much as does the discourse referent introduced in the processing of the universal quantifier (corresponding to men who own donkeys). Hence the universal force of the indefinite in the relative clause donkey (male donkey owners must beat *every* donkey they own for the sentence to be true). Second, note how the treatment of anaphora is implemented. In applying the rule for indefinites, an indefinite introduces a discourse referent; and the pronoun rule substitutes a "suitable" discourse referent for the pronoun. Often, of course, this suitable discourse referent will have been introduced by an indefinite. In such cases, the pronoun and indefinite end up being the same discourse referent/free variable; and the default existential generalization of free variables results in their being "semantically bound."

A problem with this early version of a DR approach is that it runs squarely into the proportion problem, discussed earlier. When we consider sentences like

15a. If a man owns a donkey, he usually beats it.
15b. Most men who own a donkey beat it.

straightforward extensions of the central ideas of the DR approach we are considering make both sentences true in a case in which one man owns 100 donkeys and beats them all, while 99 other men each own 1 donkey and don't beat them. The problem is that on the DR approach, 'usually' in 15a quantifies over farmer/donkey pairs; and 'Most' in 15b does as well (recall how in the relative clause donkey sentence considered above, 'Every' bound both the free variables/discourse referents for farmers and donkeys). Though this problem was corrected in Kamp and Reyle (1993), the proportion problem undermines what was thought to be an insight of DR approaches: namely the idea that the quantificational force of indefinites is inherited from other quantifiers that bind multiple variables. A second issue with the simple DR approach discussed to this point is that anaphora on indefinites is understood as essentially binding: As we saw, an indefinite introduces a discourse referent that is then "picked up" by a pronoun with the result that default existential generalization of free variables/discourse referents treats them both as effectively bound variables. But this means that DR approaches will have to invoke some other mechanism to explain many cases of anaphora on indefinites, such as the following:

16. A woman is following Glenn. John thinks she is a spy.

In 16 the second sentence can be read as attributing a *general* belief to John according to which he does not have a belief about any particular woman, but believes (roughly) that the woman who is following Glenn is a spy. But on DR approaches, given the mechanisms we have discussed to this point, such sentences only have readings on which it is as though the anaphoric pronoun is bound by the indefinite, and on this reading the second sentence of 16 asserts that John has a belief about a particular woman. See Asher (1987) for an attempt to extend DR approaches to capture this sort of data which results in significant additions to and complications of DR approaches. Note that other accounts (some descriptive approaches and the CDQ account) *can* get the relevant readings here straightforwardly, which would seem to be advantage of such approaches with respect to this data.

Let me turn now to *dynamic approaches*. Here, there are many versions and so I have opted to introduce the reader to the general features of such approaches with one of the simplest and most influential versions: Groenendijk and Stokhof (1991). Henceforth, I shall refer to their account as GSDL.

To begin with, let's look at how simple discourse anaphora is handled on GSDL. So consider:

17. A man loves Joanna. He is rich.

Now in GSDL, indefinites such as 'a man' are treated as existential quantifiers. Further, GSDL idealizes a bit and treats consecutive sentences in discourses as being conjoined. So we can think of 17 as follows:

17a. (∃x)(man x & x loves Joanna) & x is rich

Here we have rendered the anaphoric pronoun 'He' as the variable 'x', the same variable that is the variable of its quantifier antecedent. This represents the anaphoric connection. The important point to notice is that the anaphoric pronoun/variable in 17a is not within the syntactic scope of its quantifier antecedent. This corresponds to the fact that in GSDL, the syntactic scopes of quantifiers are confined to the sentences in which they occur, as current syntactic theory tells us they should be.

The key to understanding the GSDL account of discourse anaphora lies in understanding the semantic accounts it offers of the existential quantifier and conjunction. Let's begin with existential quantification. In standard first-order logic, we can interpret formulas (in a model) by assigning them sets of sequences of individuals (those that satisfy the formulas). GSDL just takes this a step further by assigning to formulas (in a model) sets of *pairs* of sequences, thought of as the input and output sequences. In the case of an existentially quantified formula '(∃x)Φ', the idea is that a pair of sequences <g,h> is in its interpretation just in case there is sequence k differing from g at most on x such that <k,h> is in the interpretation of 'Φ'. Note how "interpreting" the existential quantifier results in shifting from the "input" sequence g to k, where k is now the "input" sequence to 'Φ'. This makes the existential quantifier "internally dynamic," capable of affecting the interpretation of expressions within its syntactic scope. Further, the fact that the output sequence of interpreting the whole existentially quantified sentence, here h, is allowed to be a sequence different from the input to the interpretation, here g, means that the processing of the existentially quantified formula may affect the interpretation of expressions *after* the existentially quantified formula, and hence *outside* the scope of the existential quantifier. This is to say that the existential quantifier is "externally dynamic," capable of affecting the interpretation of expressions *outside* its syntactic scope. As we will see, an expression can be internally dynamic and externally static (as well as internally and externally static). At any rate, putting things very roughly, the idea here is that once the existential quantifier "resets" the value of 'x' in a sequence so that it satisfies the formula that the quantifier embeds, that value stays reset and can affect the interpretation of subsequent formulas.

Turning now to conjunction, the idea here is similar. Again, the fundamental idea is that the interpretation of the left conjunct can affect the interpretation of the right conjunct. A bit more formally, a pair of sequences <g,h> satisfies a conjunction just in case there is a sequence k such that <g,k> satisfies the left conjunct and <k,h> satisfies the right conjunct. Note how interpreting the left conjunct changes the input sequence for the interpretation of the right conjunct. Again, this means that conjunction is "internally dynamic," possibly affecting the interpretation of expressions in its scope. And again, that the output of interpreting a conjunction, here h, can differ from the input, here g, means that a conjunction is capable of affecting things outside of it and hence outside of the scope of that conjunction sign. Again, this is to say that conjunction is "externally dynamic."

Given these treatments of existential quantification and conjunction, it is easy to show that on GSDL the following two formulas are equivalent even when 'ψ' contains free occurrences of 'x':

(∃x)(ϕ) & ψ and (∃x)(ϕ & ψ)

So if we consider again our example of discourse anaphora 17 and its "representation" in GSDL 17a:

17a. $(\exists x)(\text{man } x \,\&\, x \text{ loves Joanna}) \,\&\, x \text{ is rich}$

this ends up being equivalent to

17b. $(\exists x)(\text{man } x \,\&\, x \text{ loves Joanna} \,\&\, x \text{ is rich})$

and so the sentences of the discourse are true iff some rich man loves Joanna. Since conjunction is externally dynamic, we can keep adding sentences with anaphoric pronouns to similar affect. Thus in a discourse such as

17c. A man loves Joanna. He is rich. He is famous.

the sentences are all true iff some rich famous man loves Joanna. These truth conditions seem intuitively correct.

Because the treatment of donkey anaphora is a bit more complicated technically, and because some of the main ideas of GSDL are now on the table, I will be more suggestive here. I urge the interested reader to consult Groenendijk and Stokhof [1991] directly. Further, since the treatments of conditional and relative clause donkey sentences are similar, I'll only consider one of them.

So consider conditional donkey anaphora:

13. If Sarah owns a donkey, she beats it.

This gets regimented in GSDL as follows:

13a. $(\exists x)(\phi) \rightarrow \psi$

where 'ψ' contains an occurrence of the variable 'x' ('it') not in the scope of the existential quantifier in the antecedent.

There are three crucial points to the GSDL treatment here: (1) the existential quantifier is externally dynamic and hence may affect the interpretation of variables outside its scope, and in particular 'x' in the consequent of 13a; (2) '$\rightarrow$' is internally dynamic and allows the interpretation of its antecedent to affect the interpretation of its consequent (just as is conjunction). Points 1 and 2 together mean that the quantifier in the antecedent of 13/13a can "semantically" bind the variable in the consequent, even though it is not in the syntactic scope of the quantifier. But without doing anything further, we would be left with 13a having the truth conditions of

13b. $(\exists x)(\phi \rightarrow \psi)$

where '$\rightarrow$' is the standard material conditional. This doesn't give the intuitive truth conditions of 13 on the reading that concerns us, since 13b would be true if something failed to be a donkey Sarah owns. The third and final element we need to get the truth conditions to come out right is to say that a pair of sequences <h,h> is in the interpretation of a conditional iff for all k such that <h,k> satisfies the antecedent, there is a j such that <k,j> satisfies the consequent. (That we only look at pairs <h,h> means that

conditionals are "externally static," and so cannot affect the interpretation of expressions outside of them, unlike conjunctions. This explains the following contrast: 'A man came in and he was happy. He was rich.'; *'If a man came in, he was happy. He was rich.') This says, roughly, that for any output sequence k of a pair of sequences satisfying the antecedent of the conditional (assigning a donkey Sarah owns to x in the case of 13/13a), k is the input of a pair <k,j> that satisfies the consequent, for some j. In the case of a simple example like 13/13a, j=k. So that the account claims that any output of a pair of sequences that satisfies the existentially quantified antecedent (which means that the sequence assigns to 'x' a donkey Sarah owns), satisfies 'Sarah beats x', and so also assigns to 'x' something Sarah beats. That is, the truth of 13/13a requires Sarah to beat every donkey she owns.

On the other hand, GSDL and dynamic approaches generally face an issue that arose in our discussion of DR approaches. Put crudely, GSDL solves the problems of discourse and donkey anaphora by formulating semantics for quantifiers that allows quantifiers to semantically bind variables that aren't in their syntactic scopes. In this they (self-consciously) resemble DR approaches. Consider again our discourse 16, repeated here:

16. A woman is following Glenn. John thinks she is a spy.

As mentioned above, 16 has a reading on which the second sentence of the discourse attributes a general belief to John. On a dynamic approach to 16, the quantifier in the first sentence semantically binds the variable in the second sentence. But then this semantically amounts to quantification into the verb of attitude, and so will not result in a reading of the second sentence on which it attributes a general belief to Scott. Hence, dynamic approaches need to invoke some other mechanism to get the reading of the second sentence in question. There are many examples of this sort that will require dynamic approaches to invoke additional mechanisms.

References

Asher, Nicholas, 1987, 'A Typology for Attitude Verbs and Their Anaphoric Properties, *Linguistics and Philosophy*, volume 10, number 2.

Barker, Chris and Chung-chieh Shan, 2008, 'Donkey Anaphora is In-Scope Binding,' *Semantics and Pragmatics*, volume 1, Article 1: 1–46.

Brasoveanu, Adrian, 2008, 'Donkey Pluralities: Plural Information States versus Non-Atomic Individuals,' *Linguistics and Philosophy*, volume 31, number 2.

Chierchia, Gennaro, 1995, *Dynamics of Meaning*, The University of Chicago Press, Chicago and London.

Cooper, Robin, 1979, 'The Interpretation of Pronouns' in Frank Heny and Helmut Schnelle, eds., *Syntax and semantics 10: Selections from the Third Groningen Round Table*, 61–92. New York: Academic Press.

Davies, Martin, 1981, *Meaning, Quantification and Necessity*, London: Routledge and Kegan Paul.

Edelberg, Walter, 1986, 'A New Puzzle About Intentional Identity,' *Journal of Philosophical Logic*, volume 15, 1–25.

Elbourne, Paul, 2005, *Situations and Individuals*, Cambridge, MA:The MIT Press.

Evans, Gareth, 1977, 'Pronouns, Quantifiers and Relative Clauses (I),' *Canadian Journal of Philosophy*, volume 7, number 3, 467–536.

Geach, Peter, 1962, *Reference and Generality*, Ithaca, NY: Cornell University Press.

Geach, Peter, 1967, 'Intentional Identity,' *Journal of Philosophy*, volume LXIV, number 20.

Groenendijk, J. and M. Stokhof, 1991, 'Dynamic Predicate Logic,' *Linguistics and Philosophy* volume 14, 39–100.

Heim, Irene, 1982, The Semantics of Definite and Indefinite Noun Phrases, Ph.D. dissertation, University of Massachusetts.

Heim, Irene, 1990, 'E-type Pronouns and Donkey Anaphora,' *Linguistics and Philosophy*, volume 13, 137–177.

Kamp, Hans, 1981, 'A Theory of Truth and Semantic Representation,' reprinted in *Formal Semantics The Essential Readings*, B.H. Partee and P. Portner, eds., Oxford:Blackwell Publishers Ltd.

Kamp, H. and U. Reyle, 1993, *From Discourse to Logic: Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht: Kluwer Academic Publishers,

King, Jeffrey C., 1987, 'Pronouns, Descriptions and the Semantics of Discourse,' *Philosophical Studies*, volume 51, 341–363.

King, Jeffrey C., 1994, 'Anaphora and Operators,' *Philosophical Perspectives*, volume 8, 221–250.

King, Jeffrey C., 2004, 'Context Dependent Quantifiers and Donkey Anaphora', in *New Essays in the Philosophy of Language, Supplement to Canadian Journal of Philosophy, vol. 30*, M. Ezcurdia, R. Stainton and C. Viger, eds., Calgary: University of Calgary Press.

King, Jeffrey C., 2005, 'Anaphora,' *Stanford Encyclopedia of Philosophy*, Stanford: Stanford University.

Kripke, Saul, 2009, 'Presupposition and Anaphora: Remarks on the Formulation of the Projection Problem,' *Linguistic Inquiry*, volume 40, number 3, 367–386.

Lewis, David, 1975, 'Adverbs of Quantification' in *Formal Semantics of Natural Language*, E.L. Keenan, ed., Cambridge: Cambridge University Press.

Neale, Stephen, 1990, *Descriptions*, Cambridge, MA: MIT Press.

Parsons, T, 1978, Pronouns as Paraphrases, Amherst: University of Massachusetts, ms.

Partee, Barbara, 1973, 'Some Structural Analogies between Tenses and Pronouns in English,' *The Journal of Philosophy*, volume 70, number 18, 601–609.

Van der Sandt, Rob, 1992, 'Presupposition Projection as Anaphora Resolution,' *Journal of Semantics*, volume 9, 333–377.

Wilson, George, 1984, 'Pronouns and Pronominal Descriptions: A New Semantical 'Category,' *Philosophical Studies*, volume 45, 1–30.

3.7
DESCRIPTIONS

Peter Ludlow

1 What are Descriptions?

Ordinarily, when philosophers talk about descriptions, they have two kinds of expressions in mind: definite descriptions—understood to be phrases of the form 'the F' (and their equivalents in other languages), and indefinite descriptions—understood to be phrases of the form 'an F' (and their equivalents in other languages). As we will see, this way of carving up the kinds of descriptions is far too blunt. First, there are many kinds of expressions that appear to have this form but that are often argued not to be descriptions. For example, in the expression 'John is a lawyer' it is often claimed that 'a lawyer' is not a genuine description, but is rather something different—a predicate for example. Furthermore, it is arguable that there are many kinds of expressions having the surface form 'the F' but that nevertheless have semantical properties unlike "typical" descriptions. For example, there are generics (as in 'The whale is a mammal'), and plurals (as in 'The dogs are barking').

Second, it is arguable that there are many expressions having surface forms quite different from 'the F' or 'an F' that could count as being descriptions. For example, it seems quite plausible that possessives like 'my father' are descriptions (as in 'the father of me'). Russell (1905) also proposed that some ordinary proper names could be construed as definite descriptions in disguise. Thus a name like 'Aristotle' might be taken as shorthand for 'the student of Plato who taught Alexander, wrote *The Nichomachean Ethics* etc.' Furthermore, as we will see, it has been argued that pronouns like 'it' might "stand proxy" for descriptions. So, for example, the pronoun in (1) might be taken to stand proxy for the corresponding definite description in (1').

(1) A man came in the room. He turned on the TV.
(1') A man came in the room. [The man who came in the room] turned on the TV.

Correlatively, it has been suggested that temporal anaphors like 'then' and modal anaphors like 'that' (in 'that would have been unfortunate') are kinds of descriptions.

If matters weren't complicated enough, recent work in philosophy and linguistics has called into question the very idea of whether the determiners 'the' and 'a' make the semantical contribution that philosophers often suppose. The problem is that only a small subset of natural languages have elements like the articles 'the' and 'a', and even for many languages that do have these elements (like English), it may well be that they are grammatical "particles" with no semantic contribution of their own. Indeed, some

recent analyses of descriptions have advanced the view that what contributions the determiners 'the' and 'a' make are not semantical, but rather pragmatic in nature. All of these possibilities will be discussed in due course, but for now we will begin with the analysis of 'the F' and 'an F' first taken up in Russell (1905, 1919).

2 Russell's Theory of Descriptions

The key idea of Russell's proposal is that a sentence like (2) containing an indefinite description, is understood to have the logical form in (2'),

(2) An F is G
(2') $\exists x(F(x) \& G(x))$

and a sentence like (3) containing a definite description is understood to have the logical form in (3').

(3) The F is G
(3') $\exists x(F(x) \& \forall y(F(y) \Rightarrow x=y) \& G(x))$

Boiled down to its simplest nontechnical form, the idea is that an expression of the form in (3) is shorthand for three claims:

i) There is an F.
ii) At most one thing is F.
iii) Something that is F is G.

Following Neale (1990) we will find it useful to substitute (iii') for (iii), which retains Russell's truth conditions and allows us to extend the theory to plural descriptions in a natural way.

iii') Everything that is F is G.

Thus tweaked, Russell's analysis is that the use of a definite description in a sentence involves an existence claim, a uniqueness claim, and a maximality claim.

3 Motivations for the Theory of Descriptions

There are three main motivations for the theory of descriptions; the first is metaphysical, the second involves semantical concerns in the philosophy of language, and the third is epistemological. Let's start with the metaphysical motivation.

Consider a negative existential sentence like (4).

(4) The Present King of France does not exist

Because definite descriptions are devices of quantification on Russell's view they can enter into scope relations with other operators—in this case, for example, negation. Accordingly, there is a kind of ambiguity in (4), between the following two logical forms.

(5a) not ([the x: x is the present King of France] exists(x))
(5b) [the x: x is the present King of France] not (exists(x))

If one wants to avoid the ontological entanglements of nonexistent objects, then one is free to say that (5b) is false (since it involves quantifying over things that don't exist) but that (5a) is true (since it is not the case that there is a present King of France). What is negated in (5a) is not a claim about some particular individual, but rather a general claim about the world—in effect a claim that the world contains exactly one individual that is presently the King of France and that whoever is presently the King of France exists.

Russell also had a number of concerns that today we might call "semantical." Consider the expressions 'The Morning Star' and 'The Evening Star'. Both refer to (or at least denote) the planet Venus, but there are contexts in which it seems incorrect to say that they have the same meaning. For example, it was as an astronomical discovery that The Morning Star was identical to The Evening Star, so it would be odd to treat an expression like 'The Morning Star is The Evening Star' as merely asserting some object to be self-identical. In a similar vein, if we utter (6),

(6) George wondered whether The Morning Star is The Evening Star.

we are most likely not saying that George was curious about whether Venus was identical to itself.

Frege proposed that that the solution to this puzzle involved the introduction of senses—abstract objects that fix the referents of these expressions, each having a different cognitive significance. In the case of (6), the Fregean solution would be to say that there are different senses attached to 'The Morning Star and 'The Evening Star'. Russell held that descriptions could do the work that senses are supposed to do. If we take the names in (6) as standing proxy for definite descriptions, then (6) can be unpacked as in (6').

(6') George wondered whether the star that appears in the morning is identical to the star that appears in the evening.

Russell saw that scope relations are relevant here as well. So, for example, sentences like (6) evince what are sometimes called de dicto/de re ambiguities. There are circumstances under which George has some object in particular in mind and is wondering, of that object, whether it is the star that appears in the evening. We can use (6) to report this fact as well; in this case we may think of the description as taking wide scope relative to the propositional attitude verb 'wondered', yielding a logical form like (7).

(7) the star x that appears in the morning is such that George wondered whether x is identical to the star that appears in the evening

And of course if George has gone mad and is in fact wondering about the object's self-identity, this may be represented as in (8), where both descriptions have wide scope.

(8) the star x that appears in the morning is such that the star y that appears in the evening is such that George wondered whether x is identical to y.

Metaphysical and semantical concerns were important to Russell in his (1905) paper, but epistemological concerns were no less significant. This became particularly clear when he authored his (1910–11) paper "Knowledge by Acquaintance and Knowledge by Description." In that paper, Russell distinguished between objects that we are directly acquainted with and objects that we only know under a description. So, for example, I might know myself by acquaintance, but I know the tallest man in Iowa only under a description.

4 Objections to the Theory of Descriptions

The theory of descriptions has encountered its fair share of criticism. This criticism has ranged from contentions that Russell simply got the truth conditions wrong in important cases to nagging worries about the details of the proposal—in particular worries relating to the nature of the descriptive content. As we will see, none of these concerns have been completely ameliorated.

4.1 The Challenge to Russell's Truth Conditions

Strawson (1950) objected that Russell's theory is simply incorrect about the truth conditions of sentences like 'The present King of France is bald'. According to Russell's analysis, this sentence is false (since it contains an existence claim to the effect that there is a present King of France), but, according to Strawson, this does not conform to our intuitions about the truth of an utterance of that sentence. In Strawson's view, an utterance of the sentence in a world where there is no present King of France is neither true nor false; perhaps the sentence has a truth value gap, or perhaps it fails to express a determinate proposition (Strawson vacillated on this), but either way it does not appear to be false. Strawson held that this fact supported a referential interpretation of expressions like 'The present King of France'—in effect, an interpretation on which 'The present King of France' is a logically proper name that only has its referent as its content.

If there is no present King of France, then an utterance containing such an expression is somehow defective on Strawson's view. It is as if I looked into my desk drawer, not allowing you to see what I was looking at, and said 'that is a fine green one.' Strawson held that utterances like these do not *entail* the existence of a fine green one or the present King of France, but rather *presuppose* their existence. If the expressions fail to refer, then there is a presupposition failure and the utterance fails to have a determinate truth value. (Notice that this sort of failure is not supposed to undermine the meaningfulness of the sentences that we utter; for Strawson, sentences are meaningful in and of themselves, independently of the utterance situation. Utterances of meaningful sentences may be true or false, or, if here is a presupposition failure, they may be neither.)

Does this whole debate come down to a case of intuition swapping? Thomason (1990: 327) and Soames (1976: 169) seemed to think so, and Strawson himself (1964) also came to doubt whether the entailment vs. presupposition debate could be settled by "brisk little formal argument[s]." Neale (1990) maintained that the matter could be settled in Russell's favor, and supported the claim by collecting a number of previously observed cases in which intuitions about truth conditions clearly do not support Strawson's view. For example, 'My mother is dating the present King of France' seems clearly false, as does 'The present King of France cleans my swimming pool', and he concluded

that these are clearly cases where the Strawsonian truth conditions have gone awry. But see von Fintel (2004) for a Strawsonian account of these intutions.

4.2 Donnellan's Distinction and the Argument from Misdescription

Donnellan (1966) observed that there is a sense in which Strawson and Russell are both right (and both wrong) about the proper analysis of descriptions. He argued that definite descriptions can be used in (at least) two different ways. On a so-called attributive use, a sentence of the form 'The F is G' is used to express a proposition equivalent to 'Whatever is uniquely F is G'. For example, on seeing murder victim Smith's badly mutilated corpse, Detective Brown might say "The murderer of Smith is insane" thereby communicating the thought that some unique individual murdered Smith and that whoever that individual is, he/she is insane. Alternatively, on a referential use, a sentence of the form 'The F is G' is used to pick out a specific individual, x, and say of x that x is G. For example, suppose Jones is on trial for Smith's murder and is behaving quite strangely at the defense table. I point at Jones and say, "The murderer of Smith is insane," thereby communicating the thought that Jones is insane (whether or not Jones is the actual murderer).

Kripke (1977) responded to Donnellan by arguing that the Russellian account of definite descriptions could, by itself, account for both referential and attributive uses; the difference between the two cases could be entirely a matter of pragmatics. Here is the idea: Grice showed us that there is an important distinction to be made between what one literally says by an utterance and what one intends to communicate (what one means) by that utterance. To take a famous example of Grice's, I might write a letter of recommendation for a student saying that he is very punctual and has excellent handwriting. Now what I have said is something about the student's punctuality and handwriting, but what I meant was that this is a very weak student.

In a similar vein, we could say that when I use a description referentially—say in Donnellan's courtroom case—I am literally making a general claim to the effect that there is a murderer of Smith and that he is insane, but what I mean by that utterance is that Jones is insane. That is, when I say 'The murderer of Smith insane' what I literally say is that exactly one person, x, is such that x murdered Smith and x is insane, but in that context I would succeed in communicating the singular proposition (about Jones) that Jones is insane. Kripke gave several reasons for thinking that this Gricean solution was preferable to an ambiguity thesis. One reason was a general methodological point that one should not introduce ambiguities blithely—doing so is a kind of philosophical cheat.

One of the advantages of employing the Gricean distinction between the proposition literally communicated and the proposition meant is that it offers an account for our being ambivalent about Donnellan's (1966) misdescription cases. In the courtroom case discussed above, I might say "Smith's murderer is insane," and still say something true even if the crazy man at the defense table is entirely innocent of the charges and the actual murderer, who is miles from the courtroom, is quite sane. At the same time there is some pull to say that in such a case one is saying something false too. We can say that this is a case where what we literally said was false but that what we intended to communicate—the proposition meant—was true. The two-level theory thus accounts for our conflicting intuitions.

Similarly, Hornsby (1977) gave the case of my observing the man ranting at the defense table and (me) saying, "The murderer of Smith is insane" not realizing that the man at the table is both innocent and quite sane, while the actual murderer is at large and quite insane. Again we are ambivalent about the truth of what I say, and as Neale (1990: 91–93) observed, the distinction between the proposition literally expressed and the proposition meant allows us to understand why. In this case, the proposition literally expressed is true, but what I intend to communicate is mistaken.

Unfortunately, there are cases where the two-stage theory doesn't appear to be sufficient. For example, consider a case where we are at the crime scene, and unbeknownst to Detective Brown there is not one murderer but several—suppose there were several perpetrators and they were all mad members of an evil cult. When Brown utters the sentence 'The murderer of Smith is insane' has he said something true or false? Again we are in two minds about the matter, but this time the distinction between what is literally said and what is meant is no help. Let's call this the "residue of the problem of misdescription" (although, strictly speaking, it is not in and of itself a misdescription case).

4.3 The Argument from Incompleteness

Another persistent problem for the classical Russellian theory of descriptions has been the charge that it fails to account for the problem of "incomplete descriptions" (for discussion see Donnellan 1968, Hornsby 1977, Devitt 1981: chapter 2, 2004, Wettstein 1981, Recanati 1986, Salmon 1982, Soames 1986, Neale 1990: chapter 2, and Reimer 1992). In addition, Kripke (1977), while defending Russell's theory of description against the problem of misdescription, allows that the argument from incomplete descriptions might be enough of a problem to force us to accept referential interpretations of descriptions. The worry, initially raised in Strawson (1950), is that if I say 'the table is covered with books', I do not mean to be suggesting that there is only one table in the world. Unfortunately, that seems to be precisely what the Russellian theory of descriptions is committed to. (Recall that on the Russellian analysis my utterance is shorthand for 'there is a table and only one table and every table is covered with books'.)

One strategy for dealing with this problem is that the context may provide us the means to flesh out the description. For example, perhaps descriptions can be fleshed out appropriately if we allow referential devices to be inserted into the description. The suggestion is that when we speak of the table we are implicitly specifying a spatial coordinate—in effect, we are saying 'the table over there'.

Neale (1990) has argued that whatever we may want to say about the problem of incompleteness, it is not very effective as an argument for the referential analysis of descriptions. For example, it appears that there are numerous examples involving quantified expressions that suffer the same fate as incomplete descriptions. I can say 'Everyone came to the party', not intending to mean everyone in the world. Or I might say, as Yogi Bera once did, 'Nobody goes there anymore, it's too crowded'. It certainly appears that what is going on in these cases is similar to the cases of incomplete definite descriptions and that there should be a single strategy for accounting for these different cases of "incompleteness." (But see Devitt 2004, who argues that these cases are genuinely different in kind, since descriptions are far more apt to be used referentially.)

Recent work has argued that the problem of incomplete definite descriptions can be accounted for if we pursue an appropriate theory of quantifier domain restriction. Stanley and Szabo (2000), for example, take this approach, suggesting that context can

restrict the domain of quantification. On their proposal, context can even shift within a sentence, allowing us to make sense of an utterance like 'The dogs barked at the dogs', where I mean to say that one group of dogs barked while the other group perhaps suffered in silence. (As their proposal is applicable to all quantified expressions and not just the theory of descriptions, I cannot focus on it here, except to note that it has spawned a great deal of discussion, including Bach 2000, Neale 2000, and Lepore 2004.)

It is important to note (following Szabo 2000 and Ludlow and Segal 2004) that even with a fully functional account of quantifier domain restriction there is a lingering problem here too. Let's call it the residue of the incompleteness problem. Consider cases like (9)

(9) Put the book on the book

Now clearly, the first use of 'the book' cannot have the same domain of quantification as the second use, since that would put two books in the domain of quantification and it would mean that both descriptions in the sentence are incomplete. But one wonders how valid a domain-shift analysis is here. Is there really a shift in the domain of quantification between the first utterance of 'the book' and the second utterance of that noun phrase?

5 Dissolving Descriptions

5.1 Unified Theories of Definite and Indefinite Descriptions

Given the Gricean resources discussed above, one might speculate that the distinction between definite and indefinite descriptions can be collapsed—that is, perhaps 'the' and 'a' have the same literal meaning and the only relevant distinction between them is pragmatic.

The motivation for this idea would be as follows. Very few natural languages have what we would recognize as definite and indefinite descriptions. In most Slavic languages, for example, 'the man' and 'a man' would both be expressed in the same way—in the determiner-free Russian equivalent of 'man'. Perhaps it is just our infatuation with surface grammatical form that leads us to think that English or Italian really has two different logical elements corresponding to surface forms 'the' and 'a'. Perhaps there is a single logical element with different pragmatic application conditions. That is, perhaps 'the' and 'a' have the same literal meaning—i.e., for both 'An F is G' and 'The F is G' the literal semantics is exhausted by the interpretation that [∃x: Fx](Gx) receives in a standard truth-conditional semantics. A number of linguists and philosophers have entertained this idea in recent years, including Kempson (1975), Breheney (1999), Szabo (2000), Zvolensky (1997), Ludlow and Segal (2004) and a version of the idea is at least considered in Heim (1982) and Kamp and Reyle (1993).

Obviously there are important differences between our application of the terms 'the' and 'a', but it does not follow that these differences in application are part of their semantical content. Indeed, many synonyms customarily are put to different uses. Take, for example, Grice (1961, 1975) on the distinction between 'but' and 'and'. On Grice's account 'but' and 'and' literally mean the same thing, but different "conventional implicatures" are associated with them; 'but' implicates a sense of contrast between the conjuncts.

Ludlow and Segal (2004) offer a similar story about 'a' and 'the'. Following a standard assumption in traditional grammar, they argue that 'the' signals that the object under discussion is given or familiar in the conversational context. Noun phrases fronted by the determiner 'a' signal that they involve new information. The idea advanced by Ludlow and Segal, however, is that this slender bit of information, combined with Gricean principles, is sufficient to generate the uniqueness implication that is carried by a definite description. That is, an existential claim that there is an F that is G, plus a signal that this is given (familiar) information, is often enough to allow us to implicate that there is unique a F that is G.

Armed with this idea, let's return to the some of the loose threads that we left hanging in section 4 above—the residue of the misdescription problem and the residue of the incompleteness problem. As we will see, the unified treatment of definite and indefinite descriptions may provide us an entering wedge for cracking open these puzzles. Let's look at the misdescription case first.

The residue of the problem of misdescription

Recall the case in which Detective Brown says 'the murderer of Smith is insane', incorrectly believing that there was one murderer when in fact the crime was committed by a ghoulish cult of insane individuals. There is a sense in which Brown spoke falsely, but there is also clearly some pull for us to say that what he said was true. As we saw in section 4, the distinction between the proposition literally expressed and the proposition meant was not sufficient to account for this ambivalence on our part. But according to Ludlow and Segal (2004), if we combine this pragmatic distinction with the unified analysis of definite and indefinite descriptions, there is something we can say about this last bit of residue.

According to the unified analysis of descriptions, what Detective Brown literally expresses is not the idea that there was a unique murderer of Smith who is insane. To the contrary, he literally expresses the proposition that there is at least one murderer of Smith who is insane. By applying Gricean principles in this context we have made out that Brown intends to say that there is a unique murderer of Smith and that he is insane. We are pulled in two directions by this case because what Brown has said is literally true but what he intended to communicate was, strictly speaking, false.

The residue of the problem of uniqueness

In section 4.3 we considered cases like (9), which did not seem to yield in a natural way to the device of quantifier domain restriction.

(9) Put the book on the book

But as Szabo (2000) and Ludlow and Segal (2004) have argued, if we combine quantifier domain restriction with the unified analysis of descriptions, the problem seems more amenable to solution. The idea is the following: What one literally expresses in (9) is that the hearer should put a book on a book. Pragmatics helps us to make out that one book in particular is being spoken of (for example the book being pointed at or looked at), and where it is to be moved.

5.2 Descriptions as Predicates

At the beginning of this chapter I noted that some uses of the expressions 'an F' and 'the F' are argued to be predicational, rather than quantificational. So, for example, consider the following cases.

(10) John is a lawyer
(11) John is the mayor of Pittsburgh

It is not unusual to think of 'is a lawyer' and 'is the mayor of Pittsburgh' as predicates. In the case of (10), we intuitively are not saying that there is a lawyer such that John is identical to that lawyer. Indeed, as Williams (1983) observed, this appearance is even more pronounced if we consider cases like (12).

(12) John is not a lawyer

Example (12) does not seem to have the meaning that there is a lawyer such that John is not identical to that lawyer. It is more natural to take an utterance of (12) as denying that John has a certain property.

If this is right, then the copula 'is' is really just the 'is' of predication and semantically dispensable (significantly, it routinely does not appear in many other languages unless it is needed to carry tense or other inflectional information). Likewise, we might say that the indefinite article 'a' is semantically inert as well—a mere "grace note" in Higginbotham's (1987) terminology. In this vein, one widely held view, due to Kamp (1984) and Heim (1982), is that we could regard indefinite descriptions as expressions containing free variables which are bound by adverbs of quantification (in the sense of Lewis 1975) interpreted as existentially quantified by the model theory, or perhaps bound by implicit existential closure operators (Diesing 1992).

For linguists it is now standard to think of indefinite descriptions following the copula as always being predicational, and it is a widespread belief that definite descriptions following the copula are often predicational. Philosophers have also been attracted to this view. See, for example, Geach (1962: section 39), Wiggins (1965: 42 ff.), Kim (1970: 211 ff.), Wilson (1978), Smiley (1981) and Higginbotham (1987) among others. Does it follow that we must admit two different kinds of descriptions (quantificational and predicational)? We saw that the analysis of singular and plural descriptions can be unified. Can we perform the same unification trick here?

Fara (2001) compellingly made the case that *all* uses of definite descriptions are predicational—even descriptions in subject position. The idea is that if I say 'The present King of France is bald', I am saying something like the following: Some x is such that x is the unique present King of France and x is bald. The interesting riff here is the notion that the uniqueness condition is built into the predicate. Extending the Sharvy (1980) analysis of plural descriptions, we get the following, where F is the set of all Fs, or the mass of all substance that is F:

(13)
$\|\text{an F}\| = F;$
$\|\text{the F}\| = \{x: x \in F \ \& \ \forall y \ (y \in F \Rightarrow y = x\}$

But again we might consider collapsing the distinction between definite and indefinite descriptions here, and say that there is a single rule that interprets all descriptions—definite, indefinite, singular, plural, (possibly even mass and generic descriptions)—as predicates. When we supplement this with the rejection of the uniqueness/maximality clause we get this result:

(13') ‖the F/the Fs/an F/some Fs‖ = F

If this is right then the determiners 'a' and 'the' are both "grace notes" not just in special positions but across the board

5.3 Descriptions as Particles

What then are we to say about the determiners 'the' and 'a'? As I noted earlier, they are rare in the world's languages and are not even employed in many Indo-European languages (for example Slavic languages, with the exception of southern Slavic languages like Bulgarian). But even in languages that deploy these articles it is clear that they are not behaving as operators. Goldsmith (2009) observes that in both ancient Greek and contemporary German the definite article can appear in several places within the same noun phrase (so that, for example, saying 'the sad clown is crying would involve the equivalent of 'the clown the sad is crying' in ancient Greek). In effect, the determiner is working as a particle that is carrying syntactic information (for example case) that is helping to glue the adjective to the noun it modifies. In other instances from contemporary German, Goldsmith introduces example constructions with two determiners that can only be understood as having one quantifier, and closely related constructions with one determiner that can only be understood as having two operators. Finally, many of the apparent wide scope readings of descriptions can be accounted for as pragmatic phenomena (for example, if there is only one familiar or given satisfier of a narrow scope description there may be the illusion of wide scope).

6 Conclusion

A close study of the syntax and semantics of natural language suggests that constructions of the form 'the F' and 'an F' are not only rare in natural languages but misleading in languages like English. These constructions really don't carry the logical significance that Russell and subsequent authors have thought. However, Russell's core insight remains intact: Many apparently referential constructions are in fact quantificational, and many terms that appear to be referential are in fact predicational. What Russell didn't see was that surface grammar is even more deceptive than he realized. Elements like 'the' and 'a' do not encode quantifiers or uniqueness clauses. The task for philosophers of language *now* is figuring out what properties of natural language *do* encode that information.

Note

1 Thanks to Gary Ostertag for helpful comments.

References

Bach, K., (2000) "Quantification, Qualification, and Context: A Reply to Stanley and Szabo," *Mind and Language* 15, 262–283.

Breheny, R, (1999) Context Dependence and Procedural Meaning: The Semantics of Definites. Ph.D. thesis, Department of Linguistics, University College, London.

Devitt, M., (1981) *Designation*, New York: Columbia University Press.

—— (2004) "The Case for Referential Descriptions," in A. Bezuidenhout and M. Reimer (eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*. Oxford: Clarendon Press, 280–305.

Devitt, M., and K. Sterelny, (1999) *Language and Reality: An Introduction to the Philosophy of Language*, 2nd edition, Cambridge, MA: MIT Press.

Diesing, M., (1992) *Indefinites*, Cambridge, MA: MIT Press.

Donnellan, K. S., (1966) "Reference and Definite Descriptions," *Philosophical Review* 75, 281–304.

Donnellan, K. S., (1968). "Putting Humpty Dumpty Back Together Again," *Philosophical Review* 77, 203–215.

Donnellan, K. S., (1978) "Speaker Reference, Descriptions, and Anaphora," in P. Cole (ed.), *Syntax and Semantics 9: Pragmatics*. New York: Academic Press, 47–68.

Fara, D.G., (2001) "Descriptions as Predicates," *Philosophical Studies* 102, 1–42. Published under the name "Delia Graff." von Fintel, K., (in 2004) "Would You Believe It? The King of France is Back! (Presupposition and Truth-Value Intuitions)," in A. Bezuidenhout and M. Reimer (eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*. Oxford: Oxford University Press.

Geach, P., (1962) *Reference and Generality*, Ithaca: Cornell University Press.

Goldsmith, L., (2009) "Articles, Particles, Syntax," Manuscript, Northwestern University, Dept. of Philosophy.

Grice, H. P., (1961) "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, Suppl. Vol. 35, 121–152.

Grice, H. P., (1975) "Logic and Conversation," in J. Cole and J. Morgan (eds.), *Syntax and Semantics, Vol 3: Speech Acts*, New York: Academic Press.

Heim, I., (1982) The Semantics of Definite and Indefinite Noun Phrases, Ph.D. Thesis, University of Massachusetts, Amherst.

—— (1990) "E-Type pronouns and Donkey Anaphora," *Linguistics and Philosophy* 13, 137–178.

Higginbotham, J., (1987) "Indefiniteness and Predication," in E. Reuland and A. ter Meulen (eds.), *The Representation of (In)definiteness*, Cambridge, MA: MIT Press, 43–70.

Hornsby, J., (1977) "Singular Terms in Contexts of Propositional Attitude," *Mind* 86, 31–48.

Kamp, H., (1984) "A Theory of Truth and Semantic Interpretation," in J. Groenendijk et al., (eds.), *Formal Methods in the Study of Natural Language*, Amsterdam Centre. Reprinted in J. Groenendijk et al. (eds.), *Truth, Interpretation, and Information*, Dordrecht: Foris, 1–43.

Kamp, H., and U. Reyle, (1993) *From Discourse to Logic*, Dordrecht: Kluwer Academic Publishers.

Kempson, R., (1975). *Presupposition and The Delimitation of Semantics*. Cambridge: Cambridge University Press.

Kim, J., (1970) "Events and Their Descriptions: Some Considerations," in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*. Dordrecht: D. Reidel, 199–215.

Kripke, S., (1977) "Speaker Reference and Semantic Reference," in P.E. French, T.E. Uehling, and H.K.Wettstein (eds.),*Contemporary Perspectives in the Philosophy of Language*, Minneapolis: University of Minnestoa Press, 6–27.

Lepore, E., (2004) "An Abuse of Context in Semantics: The Case of Incomplete Definite Descriptions," in A. Bezuidenhout and M. Reimer (eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*. Oxford: Oxford University Press.

Lewis, D., (1975) "Adverbs of Quantification," in E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press, 3–15.

Ludlow, P. and G. Segal, (2004) "On a Unitary Semantical Analysis for Definite and Indefinite Descriptions," in A. Bezuidenhout and M. Reimer (eds.) *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*. Oxford: Oxford University Press.

Neale, S., (1990) *Descriptions*, Cambridge, MA: MIT Press.

—— (2000) "On Being Explicit: Comments on Stanley and Szabo, and on Bach," *Mind and Language* 15, 284–294.

Recanati, F., (1986) "Contextual Dependence and Definite Descriptions," *Proceedings of the Aristotelean Society* 87, 57–73.
Reimer, M., (1992) "Incomplete Descriptions," *Erkenntnis* 37, 347–363.
Russell, B., (1905) "On Denoting," *Mind* 14, 479–493.
—— (1910–11) "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotelean Society*. Reprinted in *Mysticism and Logic*, London: George Allen and Unwin, 1917, and New York: Doubleday, 1957.
—— (1919) *Introduction to Mathematical Philosophy*, London: George Allen and Unwin.
—— (1957) "Mr. Strawson on Referring," *Mind* 66, 385–389.
Salmon, N., (1982) "Assertion and Incomplete Descriptions," *Philosophical Studies* 42, 37–45.
Sharvy, Richard (1980) "A More General Theory of Definite Descriptions," *The Philosophical Review* 89, 607–623.
Smiley, T.J., (1981) "The Theory of Descriptions," *Proceedings of the British Academy* 67, 331–337.
Soames, S., (1976) A Critical Examination of Frege's Theory of Presupposition and Contemporary Alternatives, Ph.D. thesis, MIT Dept. of Linguistics and Philosophy.
—— (1986) "Incomplete Definite Descriptions," *Notre Dame Journal of Formal Logic* 27, 349–375.
Stanley, J. and Z. Szabo, (2000) "On Quantifier Domain Restriction," *Mind and Language* 15, 219–261.
Strawson, P. F., (1950) "On Referring," *Mind*, 59, 320–334.
—— (1964) "Identifying Reference and Truth-Values," *Theoria* 3, 96–118.
Szabo, Z., (2000) "Descriptions and Uniqueness," *Philosophical Studies* 101, 29–57.
Thomason, R., (1990) "Accomodation, Meaning, and Implicature: Interdisciplinary Foundations for Pragmatics," in P. Cohen, J. Morgan, and M. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press, 325–363.
Wettstein, H., (1981) "Demonstrative Reference and Definite Descriptions," *Philosophical Studies* 40, 241–257.
Wiggins, D., (1965) "Identity Statements," in R.J.Butler (ed.), *Analytical Philosophy* 2nd edition. Oxford: Basil Blackwell.
Williams, E., (1983) "Semantic vs. Syntactic Categories," *Linguistics and Philosophy* 6, 423–446.
Wilson, G., (1978) "On Definite and Indefinite Descriptions," *The Philosophical Review* 87, 48–76.
Zvolensky, Z. (1997) "Definite Descriptions: What Frege got Right and Russell Didn't," *Aporia* 7, 1–16.

3.8
PLURALS

Bernhard Nickel

The study of plurality has implications for our understanding of the semantics of natural language and broader issues in philosophy, largely by bearing on matters logical. Perforce, I'll emphasize some issues at the expense of others, and I'll indicate directions for further study along the way. Other overviews, with different points of emphasis, are Lønning (1997), Schein (2006), and Rayo (2007).

In the first instance, plurality is a morphological notion and at least in English, it distinguishes between count and mass nouns. While the former can exhibit it *(dogs, trees, tables, chairs),* the latter cannot *(*informations, *advices, *broccolis).* The semantics of plurality are far less straightforward.

1 More Than One?

The simplest hypothesis is that plural morphology means *more than one*. That cannot be the whole story, however.

(1) a. The President met with bankers on Friday.
    b. The President met with more than one banker on Friday.
(2) a. The President didn't meet with bankers on Friday.
    b. The President didn't meet with more than one banker on Friday.
(3) a. Did you see whales on your trip?
        i. Yes, I saw (just) one.
        ii. # No, I saw (just) one.
    b. Did you see more than one whale on your trip?
        i. # Yes, I saw (just) one.
        ii. No, I saw (just) one.

While (1a) and (1b) convey exactly the same information, (2a) and (2b) do not. In the scope of negation, the plural seems to mean *one or more*. The same is true of questions, as (3a) and (3b) illustrate. Since (3(a)i) but not (3(a)ii) is a proper response to (3a), the plural seems to mean *one or more*. Example (3b) shows that if it meant *more than one,* we should expect the opposite pattern of felicitous answers. For this reason, various theorists have suggested that the plural always means *one or more,* and that the "more than one" interpretation is an implicature of some sort. Krifka (2004), Sauerland et al. (2005), and Spector (2007) implement this strategy in the context of truth-conditional semantics, Kamp and Reyle (1993) in Discourse Representation Theory.

A similar variability, with an explicit concern for compositionality, was pointed out by Chomsky (1975).

(4) That unicycle has wheels.
(5) Unicycles have wheels.

It seem as if the bare plural *wheels* means *more than one* in (4) but not in (5). The latter is what has become known as a *dependent plural* (de Mey, 1981). Chomsky argued that this presents a problem for compositionality. The plural makes different contributions to the meaning of the two sentences, but we cannot predict whether the plural means *more than one* without knowing what kind of sentence we're interpreting.

Here, too, several theorists (e.g. Spector, 2007, Zweig, 2008) have suggested that the plural is semantically number-neutral and that the appearance of a plural meaning in (4) is the result of an implicature. Working out the details of this approach may have far-reaching implications for our theory of implicature. Unlike ordinary scalar implicatures which can be canceled, we cannot cancel the "more than one" implicature here, as shown by

(6) # That unicycle has wheels, though I don't mean to suggest that it has more than one wheel.

Dependent plurals thus seem to require some other account of implicature, perhaps one on which implicatures are computed alongside (or as part of) semantic composition (Chierchia, 2002, 2006).

2 Collectivity, Distributivity, and Cumulativity

There is yet more variability to account for.

(7) All of the children slept.
(8) All of the children gathered in the yard.
(9) Three children ate four pizzas.

The predicate in (7) is what is usually called *distributive:* if it can be truly applied to a plurality, it can also be truly applied to each of its members. This connection fails in (8), making it *collective:* even if the children gathered in the yard and Jane is one of them, Jane didn't gather in the yard. Example (9) exhibits a cumulative reading: there were three children eating pizzas and between them four pizzas were eaten.

Collectively interpreted sentences such as (8) cannot be analyzed in terms of distributively interpreted sentences. We thus need to take at least collectivity as basic. But perhaps collectivity is the only basic phenomenon, and what looks like further interpretations are simply the result of indeterminacy. On this view, (7) and (8) predicate a property of a plurality, and one way for the property to truly apply to it is for it to apply to each of its members. Compare: there are many ways for the sentence *there are potatoes in the pantry* to be true—the potatoes might be in bags or in a box, but these are just ways for a single reading to be true (see Harnish, 1991, Higginbotham, 1981, Katz, 1977).

The currently most widespread view holds that the distributive interpretation is a genuine reading (see Gillon, 1987, for a range of arguments). But what determines

whether a sentence is interpreted collectively or distributively? It could be a matter of the lexical meaning of the verb, so that the meaning of *sleep* includes the meaning-postulate that whenever it is true of a collection, it's also true of each member of the collection (Scha, 1981). But we see readings that are inexplicable on this approach (see Winter, 2000).

(10) The girls wore a dress.

On the meaning-postulate approach, we can only predict the reading of (10) on which there is a single dress for all of the girls because there is no other operator that the existential quantifier *a dress* can enter into scope relations with.

Once we posit such an operator, (10) can be interpreted as *the girls are such that, for each of them, there is a dress that she wore*. This is a *distributive* operator: It distributes the predicated property to each thing in the plurality picked out by the subject.

One could think that the operator is part of the plural noun phrase (NP), so that collectivity or distributivity is a feature of the NP (see Lakoff, 1972 and Gillon, 1987, 1990, 1992). It is more likely, however, that this distributive operator is part of the verb phrase (VP), based on examples combining distributivity and collectivity (see, e.g. Beck and Sauerland, 2000, Landmann, 2000, Lasersohn, 1995, McKay, 2006, Pietroski, 2005, Schein, 1993, Schwarzschild, 1996, Winter, 2000).

(11) The children woke up and gathered in the yard.

If the NP had to be read either collectively or distributively, one of the predicates wouldn't apply. But on the VP-analysis, (11) is essentially equivalent to (12).

(12) The children are such that each of them woke up and they gathered in the yard.

Gillon (1987, 1990, 1992) and Schwarzschild (1994, 1996) have argued that in order to capture all the semantic possibilities, the analysis needs to be more complex still. Some sentences aren't collective, but the predicate cannot be distributed all the way down to the individuals involved, either. Suppose we're buying apples. Each apple costs fifty cents, and we buy twelve. (13a) and (13b) are true in that case.

(13) a. The apples cost fifty cents.
b. The apples cost six dollars.

In (13a), the VP contains a distributive operator, while it is absent in (13b). In the right context, perhaps once we're told that the apples come prewrapped in six-packs, we can also describe the situation with (14).

(14) The apples cost three dollars.

We've already exhausted the possibilities with respect to the distributive operator. When it's present, (13a) is true while (13b) and (14) are false. When it's absent, (13b) is true and the other two false. Either way, we cannot predict the true reading of (14). We'd like a formal way of roughly capturing the paraphrase (15).

(15) The apples are such that each six-pack among them costs three dollars.

Here, the predicate is distributed to collections that in turn make up the plurality, rather than to the individuals in that plurality.

Formally, we accomplish this by allowing the distributive operator to distribute the predicate to contextually salient subclasses of the plurality. One particularly striking feature of this context-dependence is that certain subclasses are completely unavailable to ordinary speakers as targets for distribution (first pointed out by Scha, 1981). In Figure 3.8.1, we can easily group the lines together in such a way as to make (16) true.

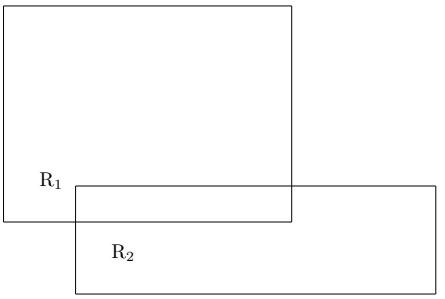


Figure 3.8.1 Rectangles

(16) The sides of R_1 run parallel to the sides of R_2.

But we cannot impose a contextual grouping on the lines in Figure 3.8.2 that verifies (17).

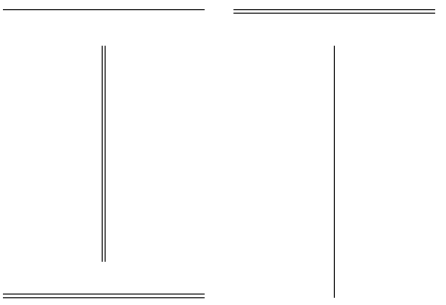


Figure 3.8.2 Lines

(17) The single lines run parallel to the double lines.

Investigating why there are such constraints promises to shed light on how context can and cannot influence interpretation. The availability of contextual restrictions on distributivity may also be relevant to other issues, such as the interpretation of definite descriptions.

2.1 Descriptions

Definite descriptions in the singular, such as *the king,* convey uniqueness, i.e., that there is only one king. Definite descriptions in the plural, such as *the children,* convey maximality, i.e., that the predicate applies to all of the children—see (18) (Sharvy, 1980, presents a theory that captures the pattern).

(18) The children slept until eight.

The uniqueness and maximality of descriptions seems to be at odds with actual usage, however. In the case of singular descriptions, this is the phenomenon of "improper descriptions" (e.g. Strawson, 1950). *The child slept* can be perfectly appropriate, even knowing that there's more than one child in the world. How to account for this is a hotly debated topic (see Cappelen and Lepore, 2005, Stanley and Szabo, 2000, for some recent debate).

The corresponding phenomenon for plural definite descriptions is that, even though descriptions seem to have a maximal flavor, they aren't perfectly equivalent to universal claims about the members of the plurality (see, e.g., Brisson, 2003, Brogaard, 2007a,b, Dowty, 1987, Scha, 1981). (18) can be true, even if some children woke up before eight. For in the case of plural definite descriptions, we have two sources of quantificational force—the description and the distributive operator—and the apparent nonmaximality of the plural description may in fact be due to a restricted distributive operator (discussed in detail by Brisson, 2003). Thus (18) might just mean (19).

(19) The children are such that each one in a (contextually determined) proper subset drawn from them slept until eight.

3 The Interpretation of "And"

Problems of interpretation also arise with a way of forming plural NPs: conjunction.

(20) a. Mary ran and Sue ran.
b. Mary and Sue ran.
c. Mary stretched and ran.

And is clearly univocal in these examples. They do not exhibit three words with distinct meanings that just happen to sound the same. But this raises a problem. When *and* connects sentences, it has a basically intersective meaning: in terms of possible worlds, the set of worlds determined by a conjunction is the intersection of the set of worlds determined by each of the conjuncts. Likewise, *and* has a basically intersective meaning when it connects predicates. The set of objects that satisfies the conjunctive predicate *tall, dark, and handsome* is the intersection of the sets that satisfy *tall* and *dark* and *handsome,* respectively. Extending this intersective idea to noun phrases works far less well. Indeed, this fact about conjunction and collectivity is one of the earliest observed and discussed facts in the philosophy of language. Thus Aristotle:

> There are two styles of [sophistical] refutation; for some depend on the language used, while some are independent of language. Those ways of producing

> the illusion which depend on language are six in number: they are homonymy, ambiguity, combination, division, accent, form of expression. [. . .] Upon division depend the proposition that 5 is 2 and 3, and even and odd, and that the greater is equal (for it is that amount and more besides).
>
> (Aristotle, 1984, 166a23–166a35)

That is, we don't want to conclude from *five is two and three* that five is both even and odd by illegitimately "dividing" the predicate into separate identity statements. The problem of explaining the logical force of this example also exercised Medieval logicians, for example Peter of Spain (1990, §75) and William of Ockham (1980, ch. 37).

In contemporary semantic theorizing, finding a unified meaning for conjunction that accounts for its basically intersective behavior in some contexts and its basically union-forming behavior in others is of interest because it ramifies into issues about how semantic composition is implemented in natural language. To say that a language (or some other system of representation) is compositional is to say that the meaning of its nonatomic parts is determined by the meaning of its atomic parts and their mode of combination. Hence, a theory that wants to describe the compositional structure of, say, English needs to say something about the semantic import of putting words together in the way that we do. One very popular idea is the following, essentially following Frege and articulated fully in Heim and Kratzer (1998). (Almost) all composition is a matter of applying functions to arguments. This basic form of function application can be used to explain the behavior of all manners of expression, including predicates of arbitrary arity, quantifiers, and adverbs. But if function application is the only mode of producing meaningful complex expressions and the surface structure of these sentences corresponds closely to their logical form, it's just about impossible to see how we could give a single meaning for *and*.

One historically popular idea for solving this problem made use of a mechanism of "conjunction reduction." We know that the surface structure of sentences is the result of transformations of an initial structure (as is shown by constraints on wh-movement, for example). Given this fact, it's at least possible that the different surface configurations in which *and* can appear are the result of a single underlying structure. Specifically, (20a) might be the basic case, and (20b) and (20c) are derived from it by "reducing the conjunction." And while this operation has syntactic and phonological effects, it doesn't have any semantic implications because the level of interpretation relevant to interpretation—the logical form—is the shared deep structure. In that case, we can make do with a single meaning for *and:* phrasal conjunction (see, e.g., Lakoff and Peters, 1969, Massey, 1976, McCawley, 1972, 1988, Smith, 1969).

Unfortunately, the conjunction reduction approach fails because it cannot account for collectively interpreted noun phrases formed using conjunctions, such as *Mary and Sue met*—it's not equivalent to *Mary met and Sue met*. For this reason, it may be required to make use of further principles of composition. The seminal paper in this area is Partee and Rooth (2002), who argue that we should allow the syntax to generate syntactic configurations that aren't interpretable by the application of function application. If we combine this with relatively tightly constrained rules about how the semantic system should deal with such failures of interpretability—for example, by applying an alternative rule in such cases—we can give a unified analysis of all occurrences of conjunction. Winter (2001) argues for a liberalized (and hence more powerful) version of this pro-

posal (an alternative view of the mechanisms of composition that focuses on conjunction rather than function application is presented in Pietroski, 2005).

4 Collectivity and Collections

So far, I've assumed that, in some sense, plural NPs pick out a plurality of things (one or more of them). I end this overview by addressing directly how best to understand this pretheoretic talk of pluralities, the issue that has been at the center of philosophers' attention. One way into the problem is to focus not on the interpretation of plurality, but on a related issue in logic and metaphysics—whether we as theorists can quantify over absolutely everything.

It is a commonplace that in ordinary speech, quantifiers are restricted. In the examples in (21), the quantifiers are not interpreted as ranging over every student that has ever existed anywhere in the world.

(21) a. Every student passed.
b. Some students didn't turn in their homework.

But for some purposes, it may be of interest to quantify over absolutely everything, for example, in making ontological claims about what sorts of things there are *not*. A metaphysician who says that there are no ghosts would not be happy if we took her to mean that there aren't any ghosts *around here*. She wants to assert something about the whole universe.

However, on a certain way of understanding our natural language, we cannot possibly quantify over everything. This issue turns on how to interpret NPs in collectively read sentences. There are basically three formal strategies. The first makes use only of standard, singular first order logic (it usually isn't emphasized that the logic is singular, but this will be important in a minute). This approach makes use of sets or some other kind of collectivizing object to capture the truth-conditions of the relevant sentences. The second retains first-order logic but adds another kind of variable, plural variables. The third makes use of the resources of second order logic. We can illustrate the three strategies by considering (22).

(22) The rocks rained down on the village.

For present purposes, we can take the whole verb phrase *rained down on the village* as an unanalyzed primitive, and I'll pretend that the only rocks in the domain of discourse are the ones that rained down on the village. The first strategy takes the expression *the rocks* to denote an object that collects the relevant rocks—perhaps the set that contains all and only those rocks, perhaps a mereological sum (Simons, 1987), or what have you. The set-based representation of (22) might then be as in (23).

(23) $\exists s(\forall x(\text{Rock}(x) \leftrightarrow x \in s)) \wedge (\text{Rained.Down}(s))$

This paraphrase has the benefit of retaining simple first order logic, a logical system that is formally and intuitively incredibly well understood. But it also has its drawbacks. A first point to make is that in a way, (23) gets the interpretation of the predicate wrong. A single object cannot rain down on anything, it can only fall, and

the situation doesn't change even if that object has parts or elements. So in order to take (23) as an interpretation of (22), we need to interpret the predicate as expressing a derived notion of "raining down on the village," call it *rain**, so that *rain** is true of an object just in case it has elements or parts and these parts rained down on the village.

Two further concerns have inclined many researchers in the area to reject the simple first order paraphrase. At least *prima facie*, it looks as if it has counterintuitive consequences about the ontological commitments we incur by speaking in a certain way. Specifically, it looks as if we as speakers of English are, when uttering (22), ontologically committed to the existence of sets in exactly the same way that we are committed to the existence of sets when we say things like (24).

(24) There is a set containing just the number one.

But intuitively, there is a real difference in the ontological commitment of a speaker who accepts only (22) and one who accepts (24). The first order logical paraphrase doesn't capture the difference since it analyzes both as quantifying over sets.

We can of course capture the difference by saying something about the theoretical machinery we use in semantics. In general, not everything we make use of in a semantic theory is something that the speaker of a natural language we interpret using that theory is committed to. For example, if we interpret quantifiers as second order relations (as in generalized quantifier theory), we don't want to say that this shows that speakers who use quantifiers in English are thereby committing themselves to the existence of sets. The use of sets is simply an artifact of efficient theorizing. We can take a similar line about the difference between (22) and (24): Take the apparent ontological commitment to sets seriously when we interpret sentences explicitly about sets, treat it as an artifact of efficient theorizing when we don't. The concern is that we wanted to use the logical paraphrase in part to lay bare when we're talking about sets in a way that leads to ontological commitments, and that isn't possible on the present line.

Finally, and this has probably been the most influential line of argument, it looks as if the first order paraphrase in terms of sets or collections can be shown to be straightforwardly mistaken in predicting that certain intuitively true sentences are incoherent. This line of argument was first presented in Boolos (1999b,c), focusing on sentences like (25).

(25) There are some things which aren't members of themselves.

Intuitively, (25) is true—every human being serves as a witness to the existential claim. But if we generally paraphrase plural noun phrases in terms of sets, (25) says that there is a set, and it contains all and only the things (including sets) that aren't members of themselves. Unfortunately, that leads directly to Russell's paradox.

A proponent of the singular first order paraphrase could block Russell's paradox by restricting the domain of quantification in a suitable way. Example (25) leads to paradox because at least in principle, anything could be something we refer to with a plural locution (especially considering the data mentioned in §1 suggesting that the plural morphology doesn't mean *more than one*). This principle is an instance of a comprehension scheme, which we might put in natural language as (26).

(26) Whenever there is one or more thing or things that have some property, then there are the things that have that property.

Now focus specifically on the consequent of (26). If plural expressions are always interpreted as denoting (possibly singleton) sets, then (26) is equivalent to (27). To make later comparisons easier, we also add the formal counterpart (28).

(27) Whenever there is one or more thing or things that have some property, then there is a set that contains all and only the things that have that property.

(28) $\exists w(\phi(w)) \rightarrow ((\exists s)(\forall y)(\mathrm{Set}(s) \wedge (\phi(y) \leftrightarrow y \in s)))$.

Now, Russell's paradox arises only if the quantifier in the consequent *all and only the things that have that property* ranges over all of the sets. To see this, imagine that we're implicitly restricting the domain of that quantificational expression. The crudest way of doing so is to exclude all sets from the domain of quantification in the object language. In that case, the original troublesome existential claim (25) can be paraphrased in the metalanguage as (29).

(29) There are some things, and they are all and only the things that *aren't sets* and aren't members of themselves.

There's obviously no commitment to the existence of a set that contains all and only the things (including all of the sets) that aren't members of themselves. Example (29) is only committed to a set containing all and only the nonsets that are also nonself-elemental. No problem there.

That we should resolve Russell's paradox by appeal to such implicit restrictions has been argued in detail by Dummett (1981, ch. 15) Dummett (1991) and Parsons (1983a, b). For concerns about the coherence of the idea that we cannot quantify over everything, see Lewis (1991) Boolos (1999a) and Williamson (2003), as well as many of the essays collected in Rayo and Uzquiano (2006).

I've just presented the problem in terms of sets and a very crude restriction on quantifiers. What is essential to the dialectic are just two parts. Plural NPs are interpreted in terms of an object that somehow collects the objects intuitively denoted by the NP and there is some kind of membership relation that relates each of these objects to this collective object—this is what Rayo (2002) calls the surrogate-strategy. Paradox is avoided by some restriction of the quantifiers of the object language. In other words, the surrogate strategy is incompatible with quantification over absolutely everything. And the surrogate strategy is the only way to capture in a singular first order language the truth-conditions of sentences containing plural NPs. Thus, if natural language "needs" to be analyzed in terms of a singular first order language, then it's impossible to quantify in natural language over everything. I'll say more about the notion of "needing to be analyzed" shortly, once we have the other options on the table.

Another way to block the derivation of Russell's paradox is to alter the formal system used to represent the semantics of the language we speak, and this comes in two flavors. The first retains a first order logic, i.e., one that only quantifies over argument positions

but allows plural variables in addition to singular ones. Thus, we allow the logical representations in (30).

(30) a. $\exists x Fx$
There is an *F*.

b. $\exists xx Fxx$
There is one or more *F*s.

The logical representation (30b) does not invoke any collectivizing entity. It allows several objects to be simultaneous arguments of a predicate. The comprehension principle (26) is formalized as in (31), where $\prec$ should be read as *is among*.

(31) $(\exists w)(\phi(w)) \rightarrow ((\exists xx)(\forall y)(\phi(y) \leftrightarrow y \prec xx))$.

When we plurally refer to one or more objects, the *xx*s in this case, we can only do so when there is at least one object among the *xx*s. Intuitively, the reason for this restriction is precisely the motivating thought for plural variables: We want to be able to refer to various objects without making use of a collectivizing entity. But because there is no such entity, it's impossible to refer to any plurality of objects without referring to some objects. In formal terms, (32), is analytically true, while (33) for example, is analytically false.

(32) $(\forall xx)(\exists y)(y \prec xx)$.
For any things, there is some thing that's among them.

(33) $(\exists xx)(\forall y)(y \prec xx \leftrightarrow y \neq y)$
There are some things, namely, the nonself-identical ones.

Given such a plural first order language, Russell's paradox is no longer derivable from the simple sentence (25), rendered as (34).

(34) $(\exists xx)(\forall y)(y \prec xx \leftrightarrow y \notin y)$

The *xx*s whose extension is defined by (34) aren't themselves a set, so that (34) does not entail that there is a set that contains all and only the non-self-elemental things.

Essentially the same effect can be achieved by altering the standard singular first order logic in another way. Rather than extending it by introducing a special kind of first-order variable—plural variables—we can move to a second-order logic. Such a logic is characterized by allowing quantification over predicate positions. Formally, for example, sentences such as (35) are well-formed, where *j* is an individual constant such as a name.

(35) $(\exists X)X(j)$.

It's hard to state the meaning of second-order logic informally without incurring unwanted commitments. One might say, for example, that (35) means that there is a property that John has. But that suggests that the second-order formalism is committed to the existence of properties we can quantify over. Perhaps the most famous worry in this direction is Quine's, who suggests that second-order logic is just set-theory in "sheep's clothing" [Quine, 1986]: (35) says neither more nor less than that there is a set that John is a member of. Using his criterion of ontological commitment (Quine, 1980),

it follows that second order logic is ontologically committed to sets. Moreover, this substantive ontological commitment disqualifies second order logic from being logic at all, rather than a substantive theory about the world, which in turn raises concerns about the viability of logicism, the project of showing that arithmetic can be reduced to logic.

Why exactly Quine's criterion is supposed to have this consequence is less than completely clear. If it turns out that an existential quantifier that binds predicate positions quantifies over sets, and ordinary (unquantified) sentences with predicates can be the true instantiations of such existentially quantified sentences, then the ordinary sentences should be committed to sets—after all, the predicate must pick out one of the things quantified over. For further discussion, see Rayo and Yablo (2001), Wright (1983).

Indeed, this discussion just recapitulates Frege's concept horse problem (Frege, 1997). Quantificational idioms in natural language tend to be object oriented, so that even when we try to explain quantification over non-object positions, we end up introducing objects that play the role of non-objects, such as sets, classes, or properties. The fact that we run into the concept horse problem in trying to elucidate the semantics of second order logic has made the following observation, due to Boolos, particularly important.

If we consider a specific family of second order logical theories, those that only allow quantification over one-place predicate positions, we are restricting ourselves to the so-called monadic second order logics. And Boolos showed that we can translate monadic second-order logic completely mechanically into first order plural logic. The key idea is to think of the one-place predicates quantified over in the second order logic as denoting all of the things that fall under them, not by denoting a set of these things, but simply denoting them plurally. Thus, we simply rephrase what looks like predication in (36a) as a partitive construction in (36b) (see Higginbotham, 2000, Hossack, 2000, Linnebo, 2004, for further discussion).

(36) a. John runs.
b. John is one of the runners.

Nonetheless, we want to keep second order logic separate from plural first order logic. As Rayo and Yablo (2001) and Williamson (2003) point out, the positions quantified over are formally distinct, one being predicative, the other objectual.

Now that the options are on the table, we can return to the issue of what it means to say that natural language "needs to be analyzed" using one or another formal tool, in this case singular first order logic, plural first order logic, or second order logic. Formal languages are artificial objects. We can use them to represent logical properties of sentences of a natural language, such as entailment, equivalence, or incompatibility because these logical relations are precisely defined in the formal language. But we also need to make sure that we understand the formal language, not by learning to speak it but by having an interpretation of the language. Such an interpretation must itself be couched in some meta-language which is primitively understood (i.e., not by being interpreted in some further meta-language). This is particularly important for the basic building blocks of the logic, its primitive vocabulary and its logical operators.

A good example concerns modality: we can have a substantive debate about whether talk about possibility or necessity is clear enough for certain philosophical purposes to simply be accepted, or whether it itself needs to be explained in terms of some other

notion. Goodman (1983) and Quine (1966) are famous skeptics about the intelligibility of our modal talk and want to reconstruct (some aspects of) that talk in other terms. In Quine (1973, 1982) are classic expressions of similar skepticism regarding second order logic. First order logic with quantification is the least controversial such logic, and the one whose basic building blocks of terms, functional expressions, relations, sentential connectives, and quantifiers have the easiest correspondence in natural language.

But as we've seen, first order logic also has its limitations. If it captures the semantics of the language we primitively understand—aka natural language—we cannot quantify over absolutely everything. So there is some incentive to conduct one's metaphysical discussions in a logic that goes beyond the strictures of singular first order logic. And the best way to show that we understand the formal tools employed in such a discussion is to show that we use basically these very same expressive resources in natural language. That is to say, if our best semantics for natural language is one that eschews the surrogate strategy in favor of plural first order or monadic second order logic, then the use of these formalisms is much less problematic. In a nice turn of phrase, Williamson suggests that the issue is what language we use as our home language, the language in which we are happy to work, at least for the time being, without seeing it through the lens of a meta-language (Williamson, 2003, 459).

4.1 Cumulativity and Events

Let's look at some data that bear on this issue. The simplest example of cumulativity is (9).

(9) Three children ate four pizzas.

The crucial observation about cumulativity is that the two quantifiers *three children* and *four pizzas* are scopeless with respect to each other. That is to say, neither of the following two scope relations captures the cumulative reading of (9).

(37) a. For each of three children *x,* there are four pizzas that *x* ate.
b. For each of four pizzas *x,* there are three children that ate *x*.

Both (37a) and (37b) are too weak, since they're true in the situation in which the cumulative reading is true, but they're also compatible with situations in which the cumulative reading is false. Example (37a) is compatible with a total of twelve pizzas being eaten by the three children, (37b) with a total of twelve children doing the eating of four pizzas. One option, proposed by Scha (1981) in a paper that shaped the debate on these issues, is that there is a special dyadic quantifier. That is, at the level of logical form, the sentence has a single element *three-four* that applies to two NPs and a relational predicate to yield the cumulative reading. If that theory is correct, we would see that natural language furnishes us with very powerful quantificational resources (formal details of binary quantifiers can be found in Keenan, 1992, van Benthem, 1989).

An alternative to the dyadic quantifier approach involves embracing a Davidsonian event-analysis for the semantics of English. The core idea of this analysis is that predicates introduce event variables into the logical forms of sentences in which they appear, and that all the other parts of the sentence are analyzed as predicating something of the

event thus introduced (see Davidson, 1980, Parsons, 1990). On this strategy, (38a) is analyzed as (38b).

(38) a. John buttered the toast.
b. $\exists e(\text{Buttering}(e) \wedge \text{Agent}(e, j) \wedge \text{Theme}(e, t))$

We can use such an event-based analysis to capture cumulative readings, as well. Informally, we analyze (9), *three children ate four pizzas* as (39).

(39) There were some eating events in which three children did the eating and four pizzas were consumed.

One very powerful reason to prefer such an event-based analysis of cumulativity to the polyadic quantificational analysis inspired by Scha comes from ditransitive verbs such as *give*. Consider (40).

(40) Three ATMs gave two customers two passwords (each).

The reading of interest is one in which we take two customers to not be distributive with respect to the three ATMs, i.e., we take there to be a total of only two customers, but where each of the customers got two passwords. This is probably the favored reading, since we can't take *two passwords* to be cumulative, as well—that would mean that somehow, two ATMs must have collaborated in the giving of at least one of the passwords.

Schein (1993) argues in great detail that this kind of mixed cumulative/distributive quantification cannot be captured with any analysis other than an event analysis along the lines of (39). On the polyadic approach, for example, we cannot scope *two passwords* "inside" the *three-two* complex that takes ATMs and customers as arguments, since there is no such spot to scope to. As such, these readings strongly speak in favor of a Davidsonian event-based analysis, though McKay (2006) has tried for an analysis that does without them.

4.2 The Unique Role Requirement

Landmann (2000), following Parsons (1990), argues that it is part of the semantics of such an event-analysis that events have a unique-role requirement. For each event, there can be only one thing that fills its agent role, one thing that fills its patient role, and so on. The kind of argument at issue concerns, for example, the analysis of (41a) as (41b).

(41) a. John and Mary met.
b. $\exists e(\text{Meeting}(e) \wedge \text{Agent}(e, j) \wedge \text{Agent}(e, m))$
c. $\exists e(\text{Meeting}(e) \wedge \text{Agent}(e, j))$
d. John met.

(41b) isn't viable as an analysis, since (41b) entails (41c), which we would think is the logical form of (41d). But clearly, (41d) isn't entailed by (41a), it isn't even grammatical.

The unique role requirement provides an argument for what I earlier called the surrogate strategy for representing collectivity. If the unique role requirement is correct, so is the surrogate strategy, since it is the only way to make it so that a single thing fills the agent role of collective predicates such as *meet, surround,* or *rain down*.

A natural response (see Pietroski, 2005, Schein, 1993) is to generalize the Davidsonian analysis to bind not a singular event variable, but a plural event variable. Using the example of cumulative readings to illustrate the idea, we can paraphrase (40) as (42).

(42) There were some events of giving something out, three ATMs did the giving, two customers were their recipients, and two passwords were given.

That is to say, all we want or can say about the ATM situation is that there were some events with the ATMs as agents that crucially do not allow us to draw any inferences regarding each of the individual events. By parallel reasoning, we analyze *John and Mary met* as (43).

(43) There were some events, they were a meeting, and John and Mary were their agents.

And though the events all together constitute a meeting, we cannot infer anything about whether the individual events that make up the meeting are themselves meetings.

This is not the end of the dialectic. Landmann cites examples such as (44a).

(44) a. The children in the choir sang a requiem.
b. There were some events, they were a singing, the children were their agents, and their theme was a requiem.

(44a) can be true if some kids were singing, some kids were just pretending, and some hid at the back of the choir and thought about comic books. And he suggests that at this point, it's not clear in what sense all of these children were agents of singing events, except insofar as they were part of a collection that, as a whole, engaged in some singing—Schein (2005, n. 62) contains a response.

5 Conclusion

Plurals provide challenges and resources for semantic theorizing. The challenge is to give a semantics that can account for distinctively plural phenomena, such as collectivity and cumulativity, while also interacting with other semantic features of a sentence, such as definite descriptions and conjunction to predict the full range of available readings compositionally. But they also allow us to motivate powerful semantic resources, such as plural quantification and additional sources of quantificational force, such as distributive quantifiers that are introduced by the verb phrase. Because plurality is such a systematic phenomenon, we can parlay empirically well-supported theorizing, such as debates about the event-based analysis of predicates, into fundamental issues about the correct logic for natural language and thence the proper logic with which to conduct our metaphysical and logical inquiry.

Related Topics

1.3 Logical Form
1.5 Implicature
1.6 Pragmatic Enrichment
2.6 Possible Worlds Semantics
2.7 Dynamic Semantics
3.4 Quantifiers and Determiners
3.5 Generics
3.7 Descriptions
3.10 Mass Terms.

References

Aristotle. Sophistical refutations. In Jonathan Barnes, editor, *The Complete Works of Aristotle*, pages 278–314. Princeton UP, Princeton, NJ, 1984.

Sigrid Beck and Uli Sauerland. Cumulation is needed: A reply to Winter (2000). *Natural Language Semantics*, 8(4): 349–371, 2000.

George Boolos. Reply to Charles Parsons' "Sets and Classes". In Richard Jeffrey, editor, *Logic, Logic, and Logic*, pages 30–6. Harvard UP, Cambridge, MA, 1999a.

George Boolos. Nominalist platonism. In Richard Jeffrey, editor, *Logic, Logic, and Logic*, pages 73–87. Harvard UP, Cambridge, MA, 1999b.

George Boolos. To be is to be a value of a variable (or to be some values of some variables). In Richard Jeffrey, editor, *Logic, Logic, and Logic*, pages 54–72. Harvard UP, Cambridge, MA, 1999c.

Christine Brisson. Plurals, *all*, and the nonuniformity of collective predication. *Linguistics and Philosophy*, 26(2): 129–184, 2003.

Berit Brogaard. Sharvy's theory of definite descriptions revisited. *Pacific Philosophical Quarterly*, 88: 160–180, 2007a.

Berit Brogaard. The but not all: A partitive account of plural definite descriptions. *Mind and Language*, 22(4): 402–426, 2007b.

Herman Cappelen and Ernie Lepore. *Insensitive Semantics*. Blackwell Publishing, Malden, MA, 2005.

Gennaro Chierchia. Scalar implicature, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti, editor, *Structures and Beyond*, pages 39–103. Oxford UP, Oxford, 2002.

Gennaro Chierchia. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, 37(4): 535–590, 2006.

Noam Chomsky. Questions of form and interpretation. In Robert Austerlitz, editor, *The Scope of American Linguistics*, pages 159–196. Peter de Ridder Press, Lisse, 1975.

Donald Davidson. The logical form of action sentences. In *Essays on Actions and Events*, pages 105–148. Oxford UP, Oxford, 1980.

Sjaak de Mey. The dependent plural and the analysis of tense. In Victoria A. Burke and James Pustejowski, editors, *Proceedings of NELS 11*. GLSA, University of Massachusetts, Amherst, MA, 1981.

David Dowty. Collective predicates, distributive predicates, and *all*. In F. Marshall, editor, *Proceedings of the Third ESCOL*. Ohio State University, Columbus, OH, 1987.

Michael Dummett. *Frege: Philosophy of Language (Second Edition)*. Harvard UP, Cambridge, MA, 1981.

Michael Dummett. *Frege: Philosophy of Mathematics*. Harvard UP, Cambridge, MA, 1991.

Gottlob Frege. On concept and object. In Michael Beaney, editor, *The Frege Reader*, pages 181–93. Blackwell, Oxford, 1997.

Brendan Gillon. The readings of plural noun phrases. *Linguistics and Philosophy*, 10(2): 199–220, 1987.

Brendan Gillon. Plural noun phrases and their readings: A reply to Lasersohn. *Linguistics and Philosophy*, 13(4): 477–485, 1990.

Brendan Gillon. Towards a common semantics for English count and mass nouns. *Linguistics and Philosophy*, 15(6): 597–640, 1992.

Nelson Goodman. *Fact, Fiction, and Forecast*. Harvard UP, Cambridge, MA, 1983.

Robert M. Harnish. Logical form and implicature. In Steven Davis, editor, *Pragmatics: A Reader*, pages 316–364. Oxford UP, New York, 1991.

Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, Malden, MA, 1998.
James P. Higginbotham. On second-order logic and natural language. In Gila Sher and Richard Tieszen, editors, *Between Logic and Intuition: Essays in Honor of Charles Parsons*, pages 79–99. Cambridge UP, Cambridge, UK, 2000.
James P. Higginbotham. Reciprocal interpretation. *Linguistic Research*, 3: 97–117, 1981.
Keith Hossack. Plurals and complexes. *British Journal for the Philosophy of Science*, 51: 411–43, 2000.
Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993.
Jerrold J. Katz. *Propositional Structure and Illocutionary Force*. Harvard UP, Cambridge, MA, 1977.
Edward Keenan. Beyond the Frege boundary. *Linguistics and Philosophy*, 15 (2): 199–221, 1992.
Manfred Krifka. Bare nps: Kind-referring, indefinites, both, or neither? In R. B. Young and Y. Zhou, editors, *Proceedings of SALT 13*. CLC Publications, Ithaca, NY, 2004.
George Lakoff. Linguistics and natural logic. In Donald Davidson and Gilbert Harman, editors, *Semantics of Natural Language*. D. Reidel, Dordrecht, 1972.
George Lakoff and Stanley Peters. Phrasal conjunction and symmetric predicates. In D. Reibel and S. Schane, editors, *Modern Studies in English: Readings in Transformational Grammar*. Prentice Hall, Englewood Cliffs, NJ, 1969.
Fred Landmann. *Events and Plurality*. Kluwer Academic, Dordrecht, 2000. Peter Lasersohn. *Plurality, Conjunction, and Events*. Kluwer, Dordrecht, 1995. David K. Lewis. *Parts of Classes*. Blackwell, Oxford, 1991.
Øystein Linnebo. Plural quantification. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2004. URL http://plato.stanford.edu/archives/win2004/entries/plural-quant/
Jan Tore Lønning. Plurals and collectivity. In Johan van Bentham and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 1009–1053. Elsevier and MIT Press, Amsterdam, 1997.
Gerald J. Massey. Tom, dick, and harry, and all the king's men. *American Philosophical Quarterly*, 13(2): 89–107, 1976.
James D. McCawley. A program for logic. In Donald Davidson and Gilbert Harman, editors, *Semantics of Natural Language*. D. Reidel, Dordrecht, 1972.
James D. McCawley. *The Syntactic Phenomena of English, Vol. I*. University of Chicago Press, Chicago, 1988.
Thomas J. McKay. *Plural Predication*. Oxford UP, Oxford, 2006.
Charles Parsons. The liar paradox. In *Mathematics in Philosophy*, pages 221–67. Cornell UP, Ithaca, NY, 1983a.
Charles Parsons. Sets and classes. In *Mathematics in Philosophy*, pages 209–20. Cornell UP, Ithaca, NY, 1983b.
Terence Parsons. *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press, Cambridge, MA, 1990.
Barbara Partee and Mats Rooth. Generalized conjunction and type amiguity. In Paul Portner and Barbara H. Partee, editors, *Formal Semantics: The Essential Readings*, pages 334–356. Blackwell, Oxford, 2002.
Peter of Spain. *Summulae logicales*. J. Benjamins Co, Amsterdam, 1990. Translated by Frances P. Dinneen.
Paul M. Pietroski. *Events and Semantic Architecture*. Oxford UP, Oxford, 2005.
W.V.O. Quine. Three grades of modal involvement. In *The Ways of Paradox and Other Essays*, pages 158–176. Harvard UP, Cambridge, MA, 1966.
W.V.O. Quine. *The Roots of Reference*. Open Court, La Salle, IL, 1973.
W.V.O. Quine. On what there is. In *From a Logical Point of View, Second Edition*, pages 1–19. Harvard UP, Cambridge, MA, 1980.
W.V.O. Quine. *Methods of Logic, Fourth Edition*. Harvard UP, Cambridge, MA, 1982.
W.V.O. Quine. *Philosophy of Logic, Second Edition*. Harvard UP, Cambridge, MA, 1986.
Agustin Rayo. Word and objects. *Noûs*, 36(3): 436–464, 2002.
Agustin Rayo. Plurals. *Philosophy Compass*, 2(3): 411–427, 2007.
Agustin Rayo and Gabriel Uzquiano. *Absolute Generality*. Oxford UP, Oxford, 2006.
Agustin Rayo and Stephen Yablo. Nominalism through denominalization. *Noûs*, 35(1): 74–92, 2001.
Uli Sauerland, Jan Anderssen, and Kazuko Yatsushiro. The plural is semantically unmarked. In S. Kesper and M. Reis, editors, *Linguistic Evidence*. Mouton de Gruyter, Berlin, 2005.
Remko Scha. Distributive, collective, and cumulative quantification. In J. Groenendijk, T. Janssen, and M. Stokhof, editors, *Truth, Interpretation, and Information (GRASS 2)*. Foris, Dordrecht, 1981.
Barry Schein. *Plurals and Events*. MIT Press, Cambridge, MA, 1993.
Barry Schein. Plurals (unabridged). ms., USC, 2005. URL http://www-rcf.usc.edu/~schein/Plurals2005.pdf
Barry Schein. Plurals. In Ernest Lepore and Barry C. Smith, editors, *The Oxford Handbook of Philosophy of*

Language, pages 716–767. Oxford UP, Oxford, 2006.
Roger Schwarzschild. Plurals, presuppositions, and the sources of distributivity. *Natural Language Semantics*, 2: 201–248, 1994.
Roger Schwarzschild. *Pluralities*. Kluwer, Dordrecht, 1996.
Richard Sharvy. A more general theory of definite descriptions. *The Philosophical Review*, 89(4): 607–624, 1980.
Peter Simons. *Parts. A Study in Ontology*. Oxford UP, Oxford, 1987.
Carlotta Smith. Ambiguous sentences with *and*. In D. Reibel and S. Schane, editors, *Modern Studies in English: Readings in Transformational Grammar*. Prentice Hall, Englewood Cliffs, NJ, 1969.
Benjamin Spector. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*, pages 243–281. Palgrave Macmillan, Houndsmills, UK, 2007.
Jason C. Stanley and Zoltan Gendler Szabo. On quantifier domain restriction. *Mind and Language*, 15(2–3): 219–261, 2000.
Peter F. Strawson. On referring. *Mind*, 59(235): 320–344, 1950.
Johan van Benthem. Polyadic quantifiers. *Linguistics and Philosophy*, 12(4): 437–464, 1989.
William of Ockham. *Summa Logicae, Part 2*. Notre Dame UP, Notre Dame, 1980. Translated by A. J. Freddoso and H. Schuurman.
Timothy Williamson. Everything. *Philosophical Perspectives*, 17(1): 415–465, 2003.
Yoad Winter. Distributivity and dependency. *Natural Language Semantics*, 8 (4): 27–69, 2000.
Yoad Winter. *Flexibility Principles in Boolean Semantics*. MIT Press, Cambridge, MA, 2001.
Crispin Wright. *Frege's Conception of Numbers as Objects*. Aberdeen UP, Aberdeen, 1983.
Eytan Zweig. Dependent Plurals and Plural Meaning. PhD thesis, New York University, 2008.

3.9
ADVERBS

Delia Graff Fara

Adverbs are a motley crew. They include words or phrases that modify (a) verbs, (b) adjectives, or (c) other adverbs:

(1) a. He fainted *surreptitiously*;
b. Their friendship was *faintly* surreptitious;
c. He fell to the ground *shockingly* quickly.

According to one contemporary grammarian, adverbs include any word or phrase that answers one of the following questions: "In what way?", "When?", "Where?", "Why?", "How?", "How much?", "Under what condition?", "To what degree?", or "How often?" (http://www.esldesk.com/grammar/adverbs (January 2011)). These diagnostic questions qualify the following words and phrases as adverbs:

(2) a. He wept *ruefully*;
b. *Yesterday*, he wept ruefully;
c. Yesterday, he wept ruefully *in his bed*;
d. Yesterday, he wept ruefully in his bed *because she lied*;
e. She managed to juggle *by using sticky tape*;
f. She managed to juggle by using *a lot of* sticky tape;
g. She managed to juggle *while drunk* by using a lot of sticky tape;
h. She managed to juggle while *very* drunk by using a lot of sticky tape;
i. She *often* managed to juggle while very drunk by using a lot of sticky tape.

(We might be skeptical about counting 'a lot' as an adverb. But otherwise the diagnostic questions seem to give no false positives.)

Words like 'therefore', which are used to relate independent clauses or states of affairs, are traditionally considered to be "conjunctive" adverbs. Here are a few more examples.

(3) a. She thought that the coffee tasted foul; *accordingly*, she spit it out.
b. She had to spit out the coffee; *otherwise* she would have retched.
c. *Between* the coffee's tasting foul and her retching after drinking it, the day got off to a very bad start.

Once we admit conjunctive adverbs to the motley crew just discussed, we can add to

our initial list of diagnostic questions the following, more cumbersome one: "What's the connection between these state of affairs?"

Adverbs can also be used to express an attitude, on the part of a speaker, toward states of affairs expressed in a complete sentence:

(4) a. Fortunately, Ian didn't fall on his head;
b. Hopefully, he'll go see a doctor.

Notice that the *sentence adverbs* in (4) would not normally be understood as falling into any of the categories of adverb described so far. For example, neither would normally be understood as modifying a verb. First, 'fortunately' in (4a) would not normally be understood as modifying 'falling'. For if I were to claim that Ian didn't fall *fortunately* on his head, I would be saying something compatible with his having fallen on his head in some other way; for example, accidentally or clumsily. But (4a) is false if Ian fell on his head *at all*—no matter how he did it. Second, 'hopefully' in (4b) would not normally be interpreted as modifying either 'go' or 'see'. If I were to claim that he will go hopefully to see a doctor or that he will go to hopefully see a doctor, I would be making a claim about the future. I would be committing myself to his going to see the doctor at some future point in time. What I would be saying, in either of those cases, would be false if he were *never* to go see the doctor. But if (4b) is truth evaluable at all, then its truth does not depend on whether he does in fact go see the doctor in the future, but only on whether *I* hope (now) that he will go see the doctor in the future.

"Sentence adverbs" need not occur at the beginning of a sentence. The following are easily interpreted as equivalent to the sentences in (4):

(5) a. Ian fortunately didn't fall on his head;
b. He'll hopefully go see a doctor.

The following sentences show, however, that although the positions of the "evaluative adverb" 'luckily' and the "epistemic adverb" 'probably' are not fixed within a sentence (6–7), their ordering relative to each other *is* fixed (8).

(6) a. Luckily, he's going to win.
b. He's luckily going to win.
(7) a. Probably, he's going to win.
b. He's probably going to win.
(8) a. Luckily, he's probably going to win.
b. # Probably, he's luckily going to win.

The syntactician Guglielmo Cinque has argued that there is a hierarchy of finely individuated categories of adverbs which constrains their ordering relative to one another, and that this holds across many languages (Cinque 1999). Cinque used such facts in turn to argue that each type of adverb had its own functional projection within a sentence, where this entails that each adverb contributes a binary-branching tree having at least three terminal nodes to the phrase structure of the sentence in which it appears. The important implication of this is that it would require syntactic structures to be significantly more complex than had been previously thought.

Let us turn to the question of how adverbial modifiers are to be represented logically. Philosophers have been concerned to capture entailments concerning adverbial modifiers as *logical* entailments. Return to some of our earlier sentences. Each sentence in (9) entails any below it.

(9) a. Alexander wept *ruefully in his bed*;
b. Alexander wept *ruefully*;
c. Alexander wept.

As an aside, note that the following sentences, in contrast to those above, will not in all contexts entail their successors:

(10) a. Alexander wept *ruefully in his bed because she lied*;
b. Alexander wept *ruefully because she lied*;
c. Alexander wept *because she lied*.

These sentences do not represent entailments since her having lied may not have been the cause of his weeping; her having lied may have been the cause of the weeping's having taken place in *his* bed as opposed, say, to her bed (in which case (9a) could be true while false). Or her having lied may have been the cause of the weeping's being *rueful* as opposed, say, to remorseful (in which case (9a) (9b)or (9c) were true while (9c) or (9b) were false). (See David Beaver and Brady Clark (2008) for a brief discussion of this feature of 'because'—in particular as it relates to *focus*—which compares the effect of focus on 'because'-clauses with the effect of focus on counterfactuals. The latter is discussed by, for example, Fred Dretske (1972), Kai von Fintel (1994), and Mats Rooth (1999).)

Leaving that issue aside, though, we now question how to logically represent the sentences in (9) so as to reflect the entailment relations that hold among them. In first-order logic, we would paraphrase 'Alexander wept' as an atomic formula, Wa, containing a monadic predicate and a name, where W stands for 'wept' and a stands for 'Alexander'. It looks, at face value, like adverbial modification can be represented in the canonical language for first-order logic in either of two ways: (i) by introducing new predicates with the same number of argument places as the unmodified original, to stand in for the modified predicate; or (ii) by adding extra argument places to accommodate the added adverbial modifiers.

To pursue the first strategy, we would paraphrase 'Alexander wept ruefully in his bed' by admitting different monadic predicates for the different ways of modifying our verb 'wept'.

(i) a. Wx: 'x wept';
b. $W'x$: 'x wept ruefully';
c. $W''x$: 'x wept ruefully in x's bed';

'Alexander wept ruefully' would be paraphrased as $W'a$.

Problem: This approach leaves the entailment relations between 'x wept', 'x wept ruefully', and 'x wept ruefully in x's bed' unaccounted for. (For discussion of this issue, see Richmond Thomason and Robert Stalnaker (1973).) The different monadic predicates W, W', et cetera, bear no logical relations to one another: No atomic monadic predication logically entails any other.

To pursue the second strategy, we would paraphrase 'Alexander wept ruefully' by admitting a two-place predicate W^2xy:

(ii) a. Wx: 'x wept';
b. W^2xy: 'x wept in way y'

'Alexander wept ruefully' would now be paraphrased as W^2ar, where r is treated as standing for *the rueful way*.

But this method of accommodating adverbial modification does not yield the desired logical entailment from 'Alexander wept ruefully' (W^2ar) to 'Alexander wept' (Wa): no atomic dyadic predication logically entails any atomic monadic one.

To fix this problem, while only slightly varying the theme, one could propose that instead of having the two different predicates for paraphrasing 'wept'—a monadic one, and a dyadic one—we could have just the dyadic one and represent the superficially monadic English predication 'x wept' as an existential quantification: 'x wept *in some way or other*' ($\exists yWxy$). This latter sentence would count as the logical paraphrase of both 'Alexander wept' and 'Alexander wept *in some way*'. The entailment of 'Alexander wept' by 'Alexander wept ruefully' would then be a special case of existential generalization, since $\exists yW^2ay$ is entailed by W^2ar.

(ii′) a. $\exists yW^2ay$: 'Alexander wept';
b. W^2ar: 'Alexander wept ruefully';

Problem: indefinitely many modifiers may be added to the monadic predicate 'wept':

(i″) a. W^3arb: 'Alexander wept ruefully in his bed';
b. W^4arbt: 'Alexander wept ruefully in his bed at two o'clock';
c. W^5arbtd: 'Alexander wept ruefully in his bed at two o'clock while dressing'.

There is no limit to the number of argument places that a single 'wept' predicate needs in order to capture, as cases of existential generalization, the entailment relations that result from adverbial modification. (See Donald Davidson 1967, 106–108.)

So the add-a-predicate approach obliterates the syntactic and semantic complexity of adverbial modification. And while the second version of the add-a-place approach does represent adverbial modification as yielding a difference in both syntactic and semantic complexity, it doesn't seem to adequately capture the hierarchical structures of those complexities. On the syntactic side, the sentence 'Alexander wept ruefully' gets paraphrased as W^2ar. But then the superficially simpler 'Alexander wept' gets paraphrased as underlyingly more complex: $\exists yW^2ay$. On the semantic side, when we say that Alexander wept ruefully, we don't think of the name and the adverb here as having the same syntactic or semantic status. When we say that Alexander wept ruefully, we don't mean that a certain relation holds, now represented as W^2, between Alexander and *the rueful way*. Rather, we mean that the property of *being rueful* applies to the property of *weeping* to yield a property in turn—namely, the property of *weeping ruefully*—that applies to Alexander.

There is a third way of giving logical expression to adverbial modification that both keeps within the bounds of a standard first-order language and yields the entailments under discussion as logical entailments. The approach is associated with Donald

Davidson (1967, 1985, e.g.) and Terence Parsons (1990, 1995, e.g.). They propose that all action sentences are to be logically represented as existential quantifications over events, with various properties being attributed to those events. (For more detailed presentation of event-based semantics, see Barry Schein "Event Semantics," (Chapter 2.8, this volume).)

For example, the past-tense sentence 'Alexander wept' is on this view to be understood as having a more complex logical form than appears on the surface. Parsons understands it as saying that there is an event that has Alexander as its agent, that is a weeping event, and that has "culminated" (since it is in the past perfect):

(11) a. Alexander wept;
b. $\exists e(\textsc{Agent}(a, e) \wedge \textsc{Weeping}(e) \wedge \textsc{Culminated}(e))$.

When we further modify 'wept' by describing the manner of the weeping, this is construed as an additional property of the event: This event was not only a weeping but also a rueful event:

(12) a. Alexander wept ruefully;
b. $\exists e(\textsc{Agent}(a, e) \wedge \textsc{Weeping}(e) \wedge \textsc{Cul}(e) \wedge \textsc{Rueful}(e))$.

Further modifications are likewise represented as attributions of properties to an event:

(13 a. Yesterday, Alexander wept ruefully in his bed;
b. $\exists e(\textsc{Yesterday}(e) \wedge \textsc{Agent}(a, e) \wedge \textsc{Weeping}(e) \wedge \textsc{Cul}(e) \wedge \textsc{Rueful}(e) \wedge \textsc{Located}(e, a\text{'s bed})$.

Given these "logical forms" (as Davidson and Parsons call them), the entailment of (11) by (12), and of (12) in turn by (13), becomes a straightforward logical entailment: an instance of *conjunction elimination* within the scope of an existential quantifier. Many points deserve mention. I will single out a few.

One might question whether the logical forms in (11–13) contain the correct predications of the weeping event in question. For example, one might question whether Alexander was really an agent of the weeping. To say that he was the agent is to say that the weeping was among his actions—that weeping was something he *did*. So shouldn't we say that he was not the agent of the weeping (unless he did it on purpose) but rather the *experiencer* of the weeping? Questions such as these don't constitute a criticism of the theory itself, just an internal criticism of one particular way of applying the theory to one particular sentence.

A similar question arises about the appropriateness of a certain other attribution to the event which does constitute an objection to the very heart of the event analysis. One might question whether the thing that was rueful was the event itself, as the analysis requires. The following illustrate some appropriate ways of attributing properties to events. Events can last for a long or short time. Events can involve knives or icepicks. But are they the kind of things that can themselves be rueful? *Being a weeping* is a property of the event, just as *taking a long time* is. But *being rueful* is not like that. It is not a property of the event. To say that an event itself was rueful is to make a category mistake. Doesn't it seem rather that *being rueful* is a property of *weeping* rather than a property of the event itself?

The event analysis incurs some ontological commitments. First, and most obviously, it requires that events *exist*—that they are in the domain of our quantifiers. Jonathan Bennett, in true purist style, points out that the theory itself is not committed to the existence of events, only to *speakers* often committing themselves to the existence of events. Given the plausible principle that speakers in the relevant cases often speak truly, the stronger commitment follows (Bennett 1988, p. 173). Second, it requires (given plausible principles) that single events may have many of their properties contingently and may correctly be described in a number of ways. Brutus's stabbing of Caesar *violently, with a knife* is the very same event as some stabbing of Caesar by Brutus. Why? To be a stabbing of Caesar by Brutus is to be an event that is a stabbing that has Brutus as its agent and Caesar as its object. To be a violent stabbing of Caesar by Brutus, with a knife, is to be a stabbing that has Caesar as its object, Brutus as its agent, a knife its instrument, and violence as a manner in which it was done. Conjunction elimination yields that the violent stabbing of Caesar, with a knife, by Brutus, is itself a stabbing of Caesar. Since there was only one stabbing of Caesar, these must be the very same event.

But why think, it might be countered, that there was only one stabbing of Caesar? After all, there was the stabbing, the violent stabbing, the violent stabbing with a knife . . . . The thought continues: These are not the same events because Brutus might have stabbed Caesar gingerly, or he might have stabbed him with an ice pick. But the *violent* stabbing couldn't have been done gingerly, and the stabbing *with a knife* couldn't have been done with an ice pick.

As one would expect, the metaphysical problems of how to individuate events manifest themselves as linguistic problems about the semantics of certain expressions that express properties of events—which includes adverbs on the views of Davidson and Parsons. One concern of Davidson's is the adverb 'slowly'; he thought it had to be excluded from his analysis. His worry was that there could be circumstances in which it would be true to say that Daisy crossed the Channel slowly and others in which it would be false to say, speaking of the very same event, that Daisy crossed the Channel slowly. Upon learning that it took Daisy ten hours to cross, one could truly say that she crossed slowly. (It normally takes just three hours to cross.) Application of the event analysis to this case would yield the following formalization:

(14) $\exists e(\mathit{Crossing}(e) \wedge \mathit{Cul}(e) \wedge \mathit{Agent}(d, e) \wedge \mathit{Object}(c, e) \wedge \mathit{Slow}(e))$

But if upon subsequently learning that Daisy *crossed by swimming*, one said that Daisy crossed quickly, one would again speak truly. This would express a truth:

(15) $\exists e(\mathit{Crossing}(e) \wedge \mathit{Cul}(e) \wedge \mathit{Agent}(d, e) \wedge \mathit{Object}(c, e) \wedge \mathit{Quick}(e))$

But since there was only one crossing event and since a single event cannot be both slow and quick, Davidson concluded that that 'slowly' and 'quickly' don't really express properties of events at all—since whether they could be truly predicated of an action depended on how that action was described or conceived—and that his analysis would therefore not apply to them. Many other adverbs would presumably fall into this class: for example, 'gracefully', 'beautifully', and 'skillfully'. But if the application of the event analysis of adverbs is really that narrow, then one wonders if it should be accepted at all.

Parsons saves the analysis by proposing—surely correctly—that just as many adjectives must be interpreted relative to a comparison class (e.g., 'tall' and 'strong'), so too

must their corresponding adverbs. This had been proposed by Gilbert Harman some years earlier (Harman 1972).

Another class of adverbs that Davidson excludes from his analysis could be those that yield *opaque* contexts. One adverb Davidson focuses on is 'intentionally'. The problem is that Mary might have intentionally emptied her bath water into the river without having intentionally contaminated the river. But the water-emptying and the contamination were one and the same event. So whether the adverb 'intentionally' applies to an event depends on how the event is conceived of or described. It therefore cannot express a *property* of an event *per se*. (The example is a variant of one of Elizabeth Anscombe's (Anscombe 1957).)

A further, significant, worry about the underlying motivation of the event analysis of adverbial modification is that its "logical forms" leave no room for a direct connection between semantic composition and syntactic composition.

Richmond Thomason and Robert Stalnaker presented an alternative theory of adverbial modification on which adverbial modifiers can be treated as predicate modifiers—ones which therefore do give logical representations of English sentences that better represent their syntactic structure as well as our intuitive understanding of their semantic structure. Their account differs in another way from the event-based analysis: It does not count 'John wept' as a structural or logical consequence of 'John wept ruefully'. The inference from the latter to the former is valid because of the particular meaning that the adverb 'ruefully' has. This seems like the right view to take since 'John allegedly wept' does not entail 'John wept'. One who makes the corresponding inference does not understand the word 'allegedly'.

Their theory involves the following strategy: first develop a logical language that allows for the existence of predicate-modifying operators and which is therefore unlike standard languages for first-order logic; give that language a well-defined semantics; and then give a semantic analysis of English sentences by correlating them with the sentences of the logical language (in cases of ambiguity, with more than one) by an intuitive rather than algorithmic procedure, but directly, without further English paraphrases as intermediaries.

The crux of their extension of first-order logic is the inclusion of predicate abstracts. A predicate abstract is formed from an open formula by prefixing it with a variable-binding abstraction operator, such as $\hat{x}$. When F is a monadic predicate, and $\hat{x}$ an abstraction operator, $\hat{x}Fx$ will be a monadic predicate. Predicate modifiers can then be prefixed to these new predicates (as well as the original ones) to form predicates. Let's look at an example. If F stands for 'fainted', then these are both monadic predicates: F and $\hat{x}Fx$. *These* can be read respectively as 'fainted' and 'is an x such that x fainted. These translations suggest that the semantics for the language should yield the semantic equivalence of the two predications. Given Thomason and Stalnaker's intensional semantics, they do: the property that is true of something in a possible world just in case F is true of that thing in that possible world.

Predicate modifiers can then be prefixed to either of these monadic predicates to yield a monadic predicate. For example, let the letter σ be the adverbial modifier 'surreptitiously'. Then σF and $\sigma\hat{x}Fx$ are equivalent monadic predicates that could be read, respectively, as 'surreptitiously fainted' and 'is an x such that x surreptitiously fainted'. These predicates would both be true of a thing in a world just in case that thing surreptitiously fainted in that world.

The main feature of including predicate abstracts in a formal language with predicate modification is that it allows for the formation of complex predicates by abstraction

from non-atomic open formulas. Consequently, abstraction makes room for scope distinctions, for example, with quantifiers. For example, in addition to the atomic formula $K(a, y)$ ('a kissed y'), there is a non-atomic formula $\forall y(Kay)$ ('a kissed everyone') that embeds it. Now let δ stand for the adverb 'deliberately'. We then have the following as distinct claims:

(16) $\forall y \delta K(a, y)$
('everyone is a y such that a deliberately kissed y');
(17) $\delta \hat{x}(\forall y K(x, y))(a)$
('deliberately kiss everyone was what a did').

These two clearly come apart. You can kiss each person deliberately without its being *kissing-everyone* that you did deliberately—for one, because you might know that you've kissed everyone even though you have in fact kissed everyone.

How would this distinction be treated on the event-based approach?

(18) $\forall x \exists e(\textsc{Agent}(a, e) \wedge \textsc{Kissing}(e) \wedge \textsc{Patient}(x, e) \wedge \textsc{Manner}(e, \textit{deliberate}))$;
(19) $\exists e(\textit{Kissing}(e) \wedge \textsc{Agent}(a, e) \wedge \textsc{Patient}(\textit{everyone}, e) \wedge \textsc{Manner}(e, \textit{deliberate}))$.

The first logical form here would represent the case involving universal quantification over individuals; the second one would represent the deliberateness of the all-kissing, which which must now be treated as a temporally extended event with a collective as patient. Moreover, if we construe the entire second statement as attributing a property to the collective *everyone*, then we must say that this is not a distributive property; otherwise

(19) would entail (18).

Predicate modification must be intensional in view of cases like the following. Suppose that all eatings are french-fry eatings: 'ate something' and 'ate french-fries' have the same extension. Still, 'surreptitiously ate something' and 'surreptitiously ate french-fries' need not have the same extension: People could be surreptitious about its being fries that they eat without being surreptitious about eating tout court. They could hide the fries but not the chewing. Accordingly, on Thomason and Stalnaker's semantics for their formal language, predicates are assigned propositional functions as their semantic values—these are functions from an individual to the set of possible worlds in which that individual satisfies that predicate. Adverbial modifiers are assigned functions from propositional functions to propositional functions, with no further semantic requirement on which functions those are. For example, some modifiers will be restrictive, while others will not be. 'Deliberately did Φ' is *restrictive* in the sense that something cannot have deliberately done Φ without having done Φ. In contrast, 'allegedly' is not restrictive. Someone might be alleged to have done something that they didn't do at all. Likewise, some modifiers may be extensional ('while wearing shorts'), while others are not ('deliberately').

We now turn to a different class of adverbs, falling under a phrase coined by David Lewis—"adverbs of quantification" (1975). Lewis demonstrates that adverbs can be used not only for predicate modification but also for quantification. Like 'more than

three', they can answer the question "How many?" And like 'most', they can answer the question "What proportion?" The following pairs illustrate the equivalence between adverbs of quantification and standard quantificational expressions.

(20) How many? (A small number.)

a. Ballerinas rarely smoke.
b. Few ballerinas smoke.

(21) What proportion? (More than half.)

a. Ballerinas usually smoke.
b. Most ballerinas smoke.

The adverbs in these sentences can each be understood as quantifying over ballerinas.

There are interpretations of (20) and (21), however, in which the 'ly' words clearly do have adverbial uses—at least, that is, if phrases answering the question "How often?" are all adverbs. For these sentences have *frequency* readings—distinct from their quantificational readings—that can be equated with the following:

(20) c. Smoking is something that ballerinas *rarely* do.
⇒ (True, for example, if: ballerinas (in general) are not the agents of smoking events more than three times a year.)

(21) c. Smoking is something that ballerinas *usually* do.
⇒ (True if: ballerinas (in general) are the agents of smoking events more than thirty times a day.)

Lewis begins his discussion of adverbs of quantification with the claim that even on what seem to be their *frequency* interpretations, some adverbs really need to be interpreted as answering a "How many?" or a "What proportion?" question.

(22) Caesar always awoke before dawn.
(Lewis's example involves 'seldom' rather than 'always' (Lewis 1975, 4).)

One might conceive of this as a answer to the question, "*How often* did Caesar awake before dawn?" But, as Lewis notes, we must not interpret (22) as meaning that *all* times were times at which Caesar awoke before dawn—even if we were to restrict those to times at which Caesar was alive. Most times were times at which Caesar was not waking at all. Rather, we should interpret (22) as meaning that all of Caesar's awakenings were before-dawn awakenings. This is an answer not to a "How often?" question, but rather to a "What proportion?" question. It answers the question "What proportion of Caesar's awakenings were before-dawn awakenings?"

But still, we might respond to Lewis, the adverb in (22) is not straightforwardly an adverb of quantification: first, because it does answer a "How often?" question even if that answer can be equivalently paraphrased as an answer to a "What proportion?" question; second, because it does not quantify over any particular domain explicitly singled out in the sentence in the way that (20a) and (21a) above do.

Lewis also argued that when adverbs are used as adverbs of quantification, they need not be understood as quantifying over anything like times or even events at all.

(23) A quadratic equation usually has two solutions.

This sentence means that *most* quadratic equations have two solutions (in an appropriately defined sense of 'most'). The adverb does not quantify over times: the sentence does not mean that most times are ones at which quadratic equations have two solutions; any quadratic equation with two solutions has those very same two solutions for ever and for always. Nor, he argued, can the adverb in (23) be understood as quantifying over events—it does not mean, for example, that most *solvings* of quadratic equations result in the finding of two solutions. Most quadratic equations have never yet been and never will be solved at all.

If adverbs of quantification do not always range over times or even events, then what do they range over? Lewis's answer was that they range over what he called "cases": all cases of Caesar's awakenings were cases of before-dawn awakenings; most cases of quadratic equations are cases of equations with two solutions.

Consideration of adverb-involving variants of the "donkey sentence" ('Every farmer who owns a donkey beats it') made famous by Peter Geach (1962) led Lewis to the view that adverbs of quantification are "unselective quantifiers." They bind all free variables in their scope. Let's consider Lewis's first donkey sentence.

(24) A man who owns a donkey always beats it now and then.

Note first that there are two adverbial phrases here, 'always' and 'now and then'. The latter adverb phrase is interpreted as a frequency adverb, as providing an answer to the question "How often do the donkeys get beaten?" Its presence prevents 'always' from getting a frequency interpretation. 'Always' here must be interpreted as a universal quantifier that ranges over individuals rather than times or events. The sentence means this:

(25) *Every* man who owns a donkey is such that *every* donkey he owns gets beaten by him now and then.

It's not merely that for each man who owns a donkey, beatings by him of donkeys he owns happen now and then. Rather, it's that every donkey that is owned by a man gets beaten now and then by that man. The double occurrence of the quantifier word 'every' in (25), which represents the truth conditions of (24), leads Lewis to think that the adverb of quantification 'always' unselectively binds those variables that correspond to men as well as those that correspond to donkeys in the logical analysis of the sentence. (He does not put things this way. I am reconstructing his argument.) Lewis would offer the following paraphrase: every case of a pair <*x*, *y*>, where *x* is a man and *y* is a donkey that *x* owns, is a case in which *x* beats *y* now and then. A "case" of a pair <*x*, *y*> meeting the required conditions (of having *x* as a man and *y* as a donkey that *x* owns) is an assignment of values to the variables *x* and *y*, where those values meet the condition that *x* beats *y* now and then.

Here is another of his regimented examples.

(26) a. Sometimes it happens that *x* sells stolen goods to y, who sells them to *z* who sells them back to *x*. (Lewis 1975, 6)

The adverb of quantification 'sometimes' here ranges over all assignments to the quadruple of variables <*x*, *s*, *y*, *z*>, and the sentence says that some such assignment meets the required condition of *s*'s being stolen goods, *x*'s selling *s* to *y*, *y*'s selling *s* to *z*, and *z*'s selling *s* back to *x*. In standard logical notation this would be written as follows:

(27) $\exists s \exists x \exists y \exists z (s \text{ is stolen goods} \wedge x \text{ sells } s \text{ to } y \wedge y \text{ sells } s \text{ to } z \wedge z \text{ sells } s \text{ to } x)$.

But in order to reflect the fact that the English word 'sometimes' by itself does the duty of this string of multiple selective quantifiers, '$\exists s \exists x \exists y \exists z$', Lewis preferred to use a notation with an unselective existential quantifier, $\exists$, that binds every free variable within its scope:

(28) $\exists (s \text{ is stolen goods} \wedge x \text{ sells } s \text{ to } y \wedge y \text{ sells } s \text{ to } z \wedge z \text{ sells } s \text{ to } x)$.

An important one of Lewis's examples was an 'if' version of the donkey sentence mentioned above. That sentence, (24), contained an indefinite description with an embedded relative clause, 'a man who owns a donkey'. The relative clause can be dispensed with by re-phrasing the sentence with an 'if'-clause:

(29) If a man owns a donkey, he always beats it now and then.

Lewis observed that adverbs of quantification could be "restricted", as he put it, by an 'if' clause which would lead to what he called a "three-part construction" consisting of (i) an adverb of quantification, (ii) a restriction on that adverb, and (iii) the remaining material of the sentence. The quantifier words in these three-part constructions differ from the standard ones in two ways. They differ by being unselective—by binding all free variables in their scope. But they also differ from the standard quantifiers in being *binary* rather than *unary*. The standard quantifiers, $\forall v$, and $\exists v$, are unary in the sense that they combine with a single sentence to form a sentence, binding a selected one of the free variables in that formula.

For example, to form the first-order paraphrase,

(29′) $\forall x \forall y (x \text{ is a man and } y \text{ is a donkey and } x \text{ owns } y \rightarrow x \text{ beats } y \text{ now and then})$,

of (29), one first prefixes the selective unary quantifier $\forall y$ to a single (open) sentence, then prefixes the selective unary quantifier $\forall x$ to the resulting (open) sentence to yield (29′).

Using Lewis's unselective, binary quantifier, one forms a truth-functional equivalent of (29′) by prefixing the unselective binary quantifier $\forall$ to the open sentences: '*x* is a man and *y* is a donkey and *x* owns *y*' and '*x* beats *y* and now and then' (he uses '+' as a separator):

(30) $\forall$ + (*x* is a man $\wedge$ *y* is a donkey $\wedge$ *x* owns *y*) + (*x* beats *y* now and then).

In this case the initial clause, 'a man owns a donkey'—logically paraphrased as the open sentence '*x* is a man and *y* is a donkey and *x* owns *y*'—forms the second part of the three-part construction. The remaining material of the sentence—logically paraphrased as the open sentence '*x* beats *y* and now and then'—forms the third part of the three-part

construction. The truth conditions for the universally quantified three-part construction are given by this clause:

> $\forall$+ Φ + Ψ is true just in case all variable assignments that satisfy Φ also satisfy Ψ.

Once we associate each adverb of quantification with the explicitly quantificational expression, Q, we can generalize that truth clause:

> Q + Φ + Ψ is true just in case Q variable assignments that satisfy Φ also satisfy Ψ.

For example, the follow triples of sentences display the logical forms and truth conditions that are assigned by this procedure to the displayed English 'if'-sentences.

(31) 'Usually' $\Rightarrow$ 'Most'

a. If a man owns a donkey, it usually kicks him now and then;
b. most + (x is a man $\wedge$ y is a donkey $\wedge$ x owns y) + (y kicks x now and then);
c. Most assignments g for which $g(x)$ is a man and $g(y)$ is a donkey and $g(x)$ owns $g(y)$ are such that $g(y)$ kicks $g(x)$ now and then.

(32) 'Rarely' $\Rightarrow$ 'Hardly any'

a. If a man owns a donkey, it rarely kicks him all the time;
b. Hardly any + (x is a man $\wedge$ y is a donkey $\wedge$ x owns y) + (y kicks x all the time);
c. Hardly any assignments g for which $g(x)$ is a man and $g(y)$ is a donkey and $g(x)$ owns $g(y)$ are such that $g(y)$ kicks $g(x)$ all the time.

One could ask why the English a. sentences are assigned the logical forms displayed in the b. sentences. Irene Heim would, relatively soon after the publication of Lewis's paper, develop a systematic and elegant theory of indefinite descriptions and anaphoric pronouns that provided answers to just that question (Heim 1982). She also developed an analysis of definite descriptions within her system which allowed for a unified semantics for definites and indefinites—one which accounted for the difference between them as a difference with respect to a small number of easily statable features. In later developments by linguists, adverbs of quantification were treated as selective quantifiers that bind situation variables; the corresponding situations are subsequently claimed to have certain properties. (See, for example, Stephen Berman (1987), Heim (1990), Kai von Fintel (1994, 2004), and Paul Elbourne (2005). Angelika Kratzer (1989) gave an explicit account of the structure of the situations needed for situation-based approaches.)

There are two problems with Lewis's analysis that deserve mention before we conclude. One of these has come to be known as the proportion problem. As we already noted, adverbs of quantification can be understood as answering questions frameable as "How many?" or "What proportion?" questions. To say that children rarely like vegetables is to say the the number of children who like vegetables is small. (Or perhaps it is the proportional claim that the percentage of children who like vegetables is small.) To

say that children usually like candy is to say that the proportion of children who like candy is greater than 50 percent. Analogously, to say (33a), that a man who owns a donkey usually beats it now and then is to say (33b), that the proportion of donkey-owning men who beat their donkeys is greater than 50 percent:

(33) a. A man who owns a donkey usually beats it now and then.
b. The proportion of donkey-owning men who beat their donkeys now and then is greater than 50 percent.

But that is not quite the truth condition that Lewis's analysis assigns to (33a). Let a man-o-donkey pair be a pair consisting of a man and a donkey he owns; let a man-b-donkey pair be a pair consisting of a man and a donkey he beats now and then. Then according to Lewis's view, to say that a man who owns a donkey usually beats it now and then is to say that the proportion of man-o-donkey pairs that are man-b-donkey pairs is greater than 50 percent.

(33) a. A man who owns a donkey usually beats it now and then
c. The proportion of man-o-donkey pairs that are man-b-donkey pairs is greater than 50 percent.

The problem is, (33b) and (33c) do not have the same truth conditions. Think of it this way. We can rephrase (33b) as follows:

(33) b′. The proportion of men that are in a man-o-donkey pair that are in a man-b-donkey pair is greater than 50 percent

The difference between (33b) and (33c) shows itself when there is a single donkey-beating man who owns a disproportionate number of the world's owned donkeys. If that one man owns a thousand donkeys that he beats now and then, while the remaining five-hundred donkey-owning men, who each own one donkey, never beat their donkey *ever*, then (33c) is true while (33b)—and (33a) along with it—is false.

A second problem is that the prosody of an assertion with adverbs of quantification can affect its truth conditions. Consider a variant of (22):

(34) Caesar usually befriended his enemies.

Normally, this would be understood to mean that most of Caesar's enemies were such that he befriended them.

But the truth conditions of this sentence can also be affected by stress. Let's use capitals to indicate stress, which for our purposes is any way of making a phrase prominent. (For more detailed discussion of focal stress and its effect on meanings, see Daniel Büring "Focus and Intonation" (Chapter 1.9, this volume). Fred Dretske's "Contrastive Statements" (Dretske 1972) has been seminal in this area.)

(34′) Caesar usually befriended his ENEMIES.

With stress of 'enemies', the sentence means that most of Caesar's befriendings were befriendings of his enemies. But contrast this:

(34″) Caesar usually befriended HIS enemies.

With stress on 'his', the sentence means that most of Caesar's enemy-befriendings were befriendings of his own enemies. Further, contrast again this:

(34′) Caesar usually BEFRIENDED his enemies.

With stress on 'befriended', the sentence means that most of Caesar's actions towards his enemies were befriendings. (We would want to limit this to actions of some relevant type.) Here is another example, taken from Mats Rooth (1985, 164):

(35) In Saint Petersburg, officers always escorted BALLERINAS;
(36) In Saint Petersburg, OFFICERS always escorted ballerinas.

The first sentence means that all escortings by officers in Saint Petersburg were escortings of ballerinas. The second sentence means that all escortings of ballerinas in Saint Petersburg were escortings by officers. If Piotr, an officer, escorts Nastenka, who is not a ballerina, then (35) is false, but (36) might still be true. If Sergei, who is not an officer, escorts Larissa the ballerina, then (36) is false, but (35) might still be true.

Here we have only skimmed the surface of the issues raised by our motley crew of adverbs. I have focused on two issues discussed by philosophers: event-based semantics for adverbial modification and the semantics of adverbs of quantification. The latter topic has been taken up almost exclusively by linguists, and I have not tried to systematically set out their dialectic. There are also many strands that could be followed which we never picked up at all: for instance, how to interpret frequency adverbs, and whether we really do have an adequate characterization of the category 'adverb' on which to base our theories.

References

Anscombe, Elizabeth (1957), *Intention*, Blackwell, Oxford.
Beaver, David I. & Clark, Brady Z. (2008), *Sense and Sensitivity: How Focus Determines Meaning*, Wiley-Blackwell, Oxford.
Bennett, Jonathan (1988), *Events and Their Names*, Oxford University Press, Oxford.
Berman, Stephen R. (1987), 'Situation-based Semantics for Adverbs of Quantification', in J. Blevins & A. Vainikka, eds., *University of Massachusetts Occasional Papers*, Vol. 12, GLSA Publications, pp. 47–68.
Büring, Daniel (2011), 'Focus and Intonation', in G. Russell & D. G. Fara, eds., *The Routledge Companion to Philosophy of Language*, Routledge, London.
Cinque, Guglielmo (1999), *Adverbs and Functional Heads: A Cross-Iinguistics Perspective*, Oxford University Press, Oxford.
Davidson, Donald (1967), 'Causal Relations', *Journal of Philosophy* 64: 691–703.
Davidson, Donald (1985), 'Adverbs of Action', in B. Vermazen & M. Hintikka, eds., *Essays on Davidson: Actions and Events*, Oxford University Press, Oxford. Page references are to reprint in Davidson (2001).
Davidson, Donald (2001), *Essays on Actions and Events*, 2nd edition, Clarendon Press, Oxford.
Dretske, Fred (1972), 'Contrastive Statements', *Philosophical Review* 81(4): 411–437.
Elbourne, Paul D. (2005), *Situations and Individuals*, MIT Press, Cambridge, MA.
Geach, Peter (1962), *Reference and Generality*, Cornell University Press, Ithaca. Page references are to the 3rd edition (1980).
Harman, Gilbert (1972), 'Logical Form', *Foundations of Language* 9: 38–65.
Heim, Irene (1982), The Semantics of Definite and Indefinite Noun Phrases, PhD thesis, University of Massachusetts, Amherst.

Heim, Irene (1990), 'E-Type Pronouns and Donkey Anaphora', *Linguistics and Philosophy* 13(2): 137–177.
Kratzer, Angelika (1989), 'An Investigation of the Lumps of Thought', *Linguistics and Philosophy*, 12(5): 607–653.
Lewis, David (1975), 'Adverbs of Quantification', in E. Keenan, ed., *Formal Semantics of Natural Language*, Cambridge University Press, Cambridge, pp. 3–15.
Parsons, Terence (1990), *Events in the Semantics of English: A Study in Subatomic Sentences*, MIT Press, Cambridge, MA.
Parsons, Terence (1995), 'Thematic Relations and Arguments', *Linguistic Inquiry* 26(4): 635– 662.
Rooth, Mats (1985), Association with Focus, PhD thesis, University Massachusett, Amherst.
Rooth, Mats (1999), 'Association with Focus or Association with Presupposition?', in P. Bosch & R. van der Sandt, eds., *Focus: Linguistic, Cognitive, and Computational Perspectives*, Cambridge University Press, Cambridge, pp. 232–244.
Schein, Barry (2011), 'Event Semantics', in G. Russell & D. G. Fara, eds., *The Routledge Companion to Philosophy of Language*, Routledge, London.
Thomason, Richmond H. & Stalnaker, Robert C. (1973), 'A Semantic Theory of Adverbs', *Linguistic Inquiry* 4(2): 195–220.
von Fintel, Kai (1994), Restriction on Quantifier Domains, PhD thesis, University of Massachusetts, Amherst.
von Fintel, Kai (2004), 'A Minimal Theory of Adverbial Quantification', in H. Kamp & B. Partee, eds., *Context-Dependence in the Analysis of Linguistic Meaning*, Elsevier Ltd, Oxford, pp. 137–175.

3.10
MASS TERMS

Francis Jeffry Pelletier

1 Introduction

Mass terms are those such as 'water', 'computer software', and 'knowledge'. They are contrasted with count terms such as 'person', 'computer program', and 'belief'. Intuitively, mass terms refer to "stuff" while count terms refer to "objects." Since mass terms refer to stuff, they (but not count terms) allow for measurement: 'a liter of water', 'three CDs worth of computer software', 'many books worth of deep knowledge'. Since count terms refer to objects, they (but not mass terms) allow for counting, quantifying, and individuating: 'a person', 'three computer programs', 'each belief of his'. Philosophers from many areas have found this distinction to be of interest—the metaphysical question of the primary existence of gunk vs. things is one obvious area, but also issues in the notions of identification and reidentification have been thought to be related to the distinction: Is it the same building when all the concrete has been replaced with new concrete? Been replaced with stone? A statue can cease to exist without its constituent matter ceasing to exist. Does this mean there are two entities here: the statue and the parcel of matter? Further afield, but still relevant, are questions about the referents of "abstract" mass terms, such as '*Curiosity* is an admirable quality to have'. The fact that 'knowledge' is a mass term while 'beliefs' is a count term have led some to question the account of knowledge as a justified true belief.

The examples just given were from English. Not all languages follow English in their characterization of this distinction—indeed, perhaps no other language is quite like English. Even within the Indo-European language group, where this distinction most closely follows that of English, there are relevant differences. And in the wider realm of the world's languages, there are those that do not allow plural/singular marking on individual nouns but only on larger phrases. There are languages that do not have a plural/singular marking for noun phrases at all (nor agreement with verb phrases); there are languages that do not have quantifiers that operate on nouns or noun phrases; there are languages that do not have determiners like 'a(n)' and 'the', even though they mark singular/plural. Thus, the examples given in the previous paragraph—which make it seem that there are clear syntactic ways to distinguish count from mass nouns—do not have the same purchase (or perhaps no purchase at all) in these languages. And this can be seen as raising questions concerning the philosophical relevance of, or interest in, the distinction.

Within the study of mass terms, the philosophical problems traditionally encountered include distinguishing mass from count terms (is it a syntactic or a semantic distinction, or something else?), deciding the extent of the classification (does it include more than

noun phrases?), describing the semantic underpinnings of mass terms (since they are not true of individuals, how can a model theory be developed?), and explaining the ontology presupposed by mass terms vs. count terms. Alongside these concerns, there is the meta-philosophical question of the extent to which the linguistic practices of the speakers of a language can be used as evidence for how those speakers view reality or, indeed, as evidence for what reality is like.

The examples above suggest that the distinction is designed to subdivide noun phrases, but there have also been attempts to extend the distinction to (some) adjectives (Quine 1960: p. 104), and especially to verb phrases. Verb phrases designating processes like *to eat* would be +MASS while those designating achievements like *to prove* would be +COUNT. See Mourelatos (1978) and Bach (1982a, b), as well as later works inspired by them, for this extension of the notion +MASS/+COUNT.

2 +MASS and +COUNT as Syntax

Many descriptive grammars of English, e.g., Quirk et al. (1985), give a syntactic characterization of the +MASS/+COUNT distinction within the category of noun (N). That is, they view the fact that some noun (e.g., *water*) is a mass term as giving an explanation for why some combinations with other words are ungrammatical. For example, they might say

(1) a. Mass nouns, unlike count nouns, do not have plural forms and thus all verb agreement is singular.
b. Mass nouns, unlike count nouns, do not admit of numeral modifiers.
c. Mass nouns, unlike count nouns, do not allow "individuative" quantifiers such as *each, every, some* (stressed), *few, several, many*, . . .
d. Mass nouns, unlike singular count nouns, employ measurement terms such as *much, a lot of, (a) little*.

(There are further problematic cases for the distinction: "collectives" such as *team, committee, army, herd, alphabet*, and "pluralia tantum" which has a large number of subcategories but includes words like *pliers, binoculars, suds, intestines, bleachers, ruins, groceries, spoils, valuables, contents*, etc.)

The syntactic characterization in (1) is supposed to account for the following classifications:

(2) Mass Nouns: *water, blood, cutlery, knowledge, carpeting, advice*, . . .
(3) Count Nouns: *person, dog, spoon, belief, carpet, suggestion*, . . .

These are all Ns viewed as being in the lexicon . . . *lexical nouns*, to give them a name. The +MASS/+COUNT features are viewed by Quirk et al. and others of this syntactic persuasion to be a part of the lexical characterization of the nouns. These features are to be inherited from the lexical items into the larger and larger syntactic units that are present in extended phrases: *blood* as lexical N contains the syntactic feature +MASS, which is inherited by the common noun phrases (CNP) *bright red blood* and *bright red blood that is on the floor* and the full noun phrase (NP) *the bright red blood that is on the floor*. The fact that this longer phrase is also +MASS is what ultimately explains why

(4) *The bright red blood that is on the floor are slippery

is ungrammatical. (Because the fact that the phrase is +MASS prohibits it from being plural, as (1a) says, and hence the agreement with the verb phrase does not happen.) Violations of the constraints involving +MASS and +COUNT yield ungrammatical results that have the same status as other syntactic violations; (4) is no more a part of English than is

(5) *Dog the quickly

It might be noted that both of the mass and count categories contain terms that are "abstract": *knowledge* and *advice* are +MASS, while *belief* and *suggestion* are +COUNT.

3 +MASS and +COUNT as Semantics

Some descriptive grammars of English, e.g., Huddleston and Pullum (2002), think of the +MASS/+COUNT distinction as a description of the semantic properties of the denotation of the terms. In this type of view, mass meanings contrast with count meanings:

(6) a. Mass meanings are *true of stuff*; count meanings are *true of things*
b. Mass meanings are *divisive in their reference*; count meanings are *true of a unit as a whole*
c. Mass meanings are *cumulative in their reference*; (singular) count meanings are *not true of groups of that which they are true*
d. Stuff that mass meanings are true of *cannot be counted*; count meanings are true of *individuated items that can be counted*
e. Stuff that mass meanings are true of *can be measured*; (singular) count meanings are *not measurable*

Some theorists take the divisiveness and the cumulativity conditions together to be called the *homogeneous in reference* condition.

In the semantic view, the fundamental difference between mass and count terms is that count terms are true of *objects*—entities that are distinct from each other even while being of the same type, and thus one can distinguish and count them—while mass terms are true of *stuff* that is undifferentiated with respect to the term being used to describe it. This in turn explains why mass terms, unlike count terms, are *divisive* in their reference: they permit something that the mass term is true of to be arbitrarily subdivided and the term to be true of these parts as well. Taking the water in the glass to be something that *is water* is true of, it can be divided into parts and *is water* will be true of both parts. And again, mass terms, unlike count terms, are also *cumulative* in their reference: Putting the water contained in two glasses into a bowl yields something of which *is water* is true. But the same is not the case with a count term like *dog*. Chopping up a dog does not yield more things of which *is a dog* is true, nor do two dogs make a thing of which *is a dog* is true. (Other than in a Frankenstein-like scenario.)

In a semantic approach, the features +MASS/+COUNT are descriptions of the semantic value of lexical nouns and the larger common noun phrases, etc. Thus, they do not figure in the syntactic well-formedness constraints of a grammar, but would emerge as a description of what the semantic values of the embedded nouns are, and how these

semantic values get altered by the syntactic combination of those nouns with other words. These features do not syntactically rule anything out; the most that can be said is that certain combinations are "semantically anomalous," and hence can't be interpreted.

As with the syntactic version of +MASS/+count, the lexicon supplies individual words with a set of syntactic features and also a set of semantic values. Larger and larger phrases that contain the noun also contain the semantic information mentioned in the lexical items, modified in accordance with rules that describe the semantic effect of being syntactically combined in the manner that is employed.

The difference between +MASS/ +COUNT as syntax and +MASS/ +COUNT as semantics thus is whether these features are seen as syntactic well-formedness constraints that yield ungrammaticality when violated or as semantic interpretability constraints upon syntactically correct sentences that yield semantic anomaly when violated.

4 Some Problems for the Syntactic Approach

In the Syntactic Approach, lexical Ns are assigned either a +MASS or +COUNT feature, and this feature controls the syntactic (in)admissibility of larger phrases. But there are many words that have both mass and count meanings, for instance

(7) a. Concrete terms
  (i) a lot of chocolate / many more chocolates
  (ii) more discipline / an academic discipline
  (iii) too much paper / write a paper
 b. Abstract terms
  (i) much discussion / three different discussions
  (ii) a lot of difference / two differences
  (iii) much more data / many more data

The examples in (7) are just the tip of the iceberg, There are many more of these "dual life" terms, sometimes forming regular patterns, but sometimes not:

(8) Mass terms used "countily":
 a. Pinot Noir is *wine* / Pinot Noir is *a wine*
 b. Kim produces *sculpture* / Kim is producing *a sculpture*
 c. *Beer* on the table / Three *beers* on the table / Eight *beers* on tap
(9) Count terms used "massily"
 a. Leslie has more *car* than *garage*
 b. Chris Pronger, 6′6″ worth of *defenseman* . . .
 c. Some people like *data* better than *theory*

As (Huddleston and Pullum 2002: p. 335) remark ". . . the dual use of *chocolate* is not remotely exceptional but is representative of an extremely widespread phenomenon," and they follow this with a list of 25 examples chosen over a wide variety of types of nouns that illustrate just how widespread the phenomenon of a noun having two equally salient meanings where one is +MASS but the other +COUNT.

There's also the "universal grinder" of Pelletier (1975), which is like a meat grinder except that it can accommodate any object, no matter how large, and its teeth are so

powerful and fine that it can grind anything, no matter how strong. Insert an object that falls under any (concrete) count noun into one side . . . for example, a hat. Push the button, and the result is that there is hat all over the floor. (True despite the fact that we might have some other term, e.g., *felt*, also describing what is on the floor.) Another push of the button and we can have book all over the floor. An unfortunate accident might generate curious cat all over the floor.

One might also think of "universal packagers" in this regard, that take any item of which a mass term is true and convert it into an object. Any time there is a use for a particular type of some mass then there can be a count term that describes it—for example, *a finely silted mud*, which can be a name for a type of mud and also a predicate that is true of all individual exemplars of this type. And if there is a standardized amount of M that is employed in some use, then there will be a count term that describes this amount, such as *a beer* or *an ice cream*. Furthermore, there seems always to be a count use for any alleged mass term M, meaning (roughly) *a kind of* M. Putting all these together, a term like *a scotch* could be true of individual servings (thus being independently true of each piece of the actual matter in the various glasses), or true of the different standardized amounts (so that two instances of the same standard one-ounce serving count as only one such standardized amount), or true of the different kinds of scotch on the table or available at the bar. Thus any of 'four', 'three', or 'five' could be true answers to the question "How many scotches are on the table?"

These considerations show that the appropriate theory needs to talk about *meanings* of terms, or *uses* of the terms, or maybe *occurrences* thereof (some occurrences are +MASS, others of the same word are +COUNT). But then this is no longer a syntactic account! And the syntactic approach just doesn't work. For, it will turn out that since *any* noun can be either mass or count, a +MASS/+COUNT syntactic distinction does no work—*nothing* is ruled out by the syntactic rules.

5 Some Problems for the Semantic Approach

Many Ns have both a natural mass and a natural count sense. So the basic lexical item that gets entered into a phrase structure description of a sentence will be one of these senses. It is never very clear how this is supposed to be effected in a grammar, but we will simply assume that there is some way that this can be done. But even if we assume that the N and the correct sense is entered there nonetheless seem to be some serious difficulties that are semantic mirrors of the difficulties found in the syntactic approach.

Many formal semanticists (e.g., Link 1983; Chierchia 1998a, b; Pelletier and Schubert 1989/2003; Landman 1991) take the characteristics in (6) to be best accounted for in terms of a semi-lattice theory. A semi-lattice has no lowest elements and is atomless. The idea is that anything that *water*, for example, might be true of has subparts—things in the lattice that are its parts—of which *water* is true; and any two elements in the *water*-lattice find a joined element also in the lattice that represents the merge of those two elements.

But it should be noted that many mass terms obviously are not "atomless" in the sense required by this theory. Consider

(10) *furniture, cutlery, clothing, equipment, jewelry, crockery, silverware, footware, bedding, toast, stemware, gravel* . . .

Clearly there are atomic parts of these; yet they are considered mass terms by any of the traditional grammars. So it cannot be an atomless mereology that accounts for the mass nature of these words; and since it doesn't account for the mass nature of these particular words, there seems to be no reason to think it accounts for the mass nature of *an*y words.

Some theorists, e.g., Huddleston and Pullum (2002), take this as evidence that terms like those in (10) are of a different nature than what we have been calling 'mass terms', and are to be treated differently. Huddleston and Pullum call them 'aggregate terms' and semantically distinguish them from other mass terms by their being true of "very different sorts of things." The idea is that furniture, for example, is true of sofas, chairs, tables, carpets, and so on, and that these are "very different" from one another. But a true mass term, for example 'blood', is really true only of one kind of thing.

But one might still wonder: Are *an*y words at all that obey the condition on divisiveness? Are there really any words that are atomless—whose referent has no smallest parts? Doesn't *water* have smallest parts: H_2O molecules perhaps? A standard defense of the divisiveness condition in the face of these facts is to distinguish between "empirical facts" and "facts of language." It is an empirical fact that water has smallest parts, it is said, but English does not recognize this in its semantics: the word *water* presupposes infinite divisibility.

It is not clear that this is true, but if it is, the viewpoint suggests interesting questions about the notion of semantics. If *water* is divisive but water isn't, then water can't be the semantic value of *water* (can it?). This in turn suggests a notion of semantics that is divorced from "the world," where semantics is not be a theory of the relation between language and the world. But it also would seem not to be a relation between language and what a speaker's mental understanding is, since pretty much everyone nowadays *believes* that water has smallest parts. Thus, the mental construct that in some way corresponds to the word *water* can't be the meaning of *water* either. This illustrates a kind of tension within "natural language metaphysics" (Bach 1986a, b).

Another problem with the semantic approach to the distinction comes from pairs of words where one is mass and the other is count and yet the items in the world that they describe seem to have no obvious difference that would account for this. On the intuitive level, it seems that postulating a *semantic* difference should have some reflection in the items of reality that the terms designate. But this is just not true. There seems to be nothing in the *referent* of the following mass vs. count terms that would explain how they should be distinguished—as they intuitively are. (See McCawley 1975 for further examples.)

(11) a. Concrete terms
  (i) baklava vs. brownies
  (ii) garlic vs. onions
 b. Abstract terms
  (i) success vs. failures
  (ii) flu vs. colds

To many, these examples and their surrounding facts have seemed to prove that the linguistic features of +COUNT and +MASS do not have any backing in reality. Nor any backing in people's intuitive understanding of *when* a word will be +MASS or +COUNT or *what it is* for a word to be +MASS or +COUNT.

6 +COUNT and +MASS Semantic Theories

A rather heterogeneous group of topics has been included within the subject matter of the semantics of +MASS vs. +COUNT terms. Given that the general topic concerns the content of lexical items, we can expect that much of the discussion will differ from that of the more traditional compositional semantic theories.

The older philosophical literature on mass terms (and many other terms) tended to find different meanings for the terms under consideration, depending on what role they had in a sentence. Theorists would consider one of the following types of uses of a mass term:

- As names, as in *Water covers most of the globe*
- As predicates true of quantities/portions of matter, as in *John drank some water*
- As predicates true of objects, as in *This ring is gold*
- As predicates true of kinds/substances, as in *Claret is wine*
- As predicate modifiers, as in *She is wearing a gold bracelet*

To some this suggested that +MASS terms were ambiguous.

> In general, a mass term in predicative position may be viewed as a general term which is true of each portion of the stuff in question, excluding only the parts too small to count . . . A mass term used in subject position differs none from such singular terms as 'mama' . . . , unless the scattered stuff that it names be denied the status of a single sprawling object.
>
> (Quine 1960: pp. 97–98)

To others it suggested that a theory should pick one of these ways that mass terms are used in sentences as basic and either ignore the others or try to generate them by some "semantico-syntactic trick." Parsons (1970) took them always to be names of substances ("in the chemistry sense"), but when a mass noun M was "in predicate position" such an occurrence was transformed into *is a quantity of* M. (This same transformation occurred in some other positions also.) One of the earliest attempts to employ mereology as a tool for representing mass terms was Moravcsik (1973). In such a framework nouns (and adjectives) designated mereological wholes and the copular connection becomes *is a part of*. Pure mereological approaches have been found wanting, because of the problem of "minimal parts"—classical mereology has no minimal parts other than the empty part, yet it is generally assumed that the designations of ordinary language +MASS nouns do have minimal parts. So Moravcsik invoked a notion of *is a part having relevant structural properties*. This has also been found wanting (see Montague 1973; Pelletier 1974), but more clever ways of operating do seem to avoid these problems (e.g., Bunt 1979, 1985; see also Burge 1972; Ojeda 1993; Moltmann 1998). A parallel—and related—development has been the notion of a (join) semi-lattice to support part–whole structures. This was first urged onto the mass term semantics world by Link (1983) and picked up by many others.

A related question is whether it is possible to show a unity within the +MASS/+COUNT realm. Gillon (1992, 1999), who adopts a syntactic view of the +MASS/+COUNT distinction, argues for a "common semantics" for +MASS and +COUNT nouns. Count nouns are assigned a set of atomic individuals, and a mass noun is assigned the mereological join of

the individuals of which it is true. Chierchia (1998a, b) instead assigns mass nouns to be "inherently plural." And in Chierchia (2010), +MASS and +COUNT are given the same type of semantic value as one another, with the exception that +MASS terms, unlike +COUNT terms, are said to have "vague minimal parts."

One understudied aspect of this concerns "abstract" mass nouns (*advice, freedom, information, . . .*). Many theorists wish to invoke a mereology or a semi-lattice as the relevant semantic structure for concrete mass nouns such as *water, blood, steel, . . .*, but mereology (in particular) seems completely out of place for abstract nouns. Intuitively, +MASS is independent of whether the noun is ±ABST, and therefore any semantic technique that is correct for concrete mass terms should be applicable to abstract mass terms. Differences between abstract and concrete mass terms should be due to the ±ABST, and not due to +MASS. But mereology just doesn't apply to +ABST,+MASS terms; hence, it should not be a defining semantic feature of –abst (concrete) +MASS terms either. And so, mereology would *not* be a part of the semantics of +MASS, but at most a consequence of the interaction of a more general semantic account of +abst and +MASS. Some attempts towards an account of the semantics for +abst, +MASS terms using the semantics of comparatives can be found in Nicolas (2010); an account using the methodology of Natural Semantic Metalanguage (Wierzbicka 1996) can be found in Goddard and Wierzbicka (2010). I think all these authors would admit that their account needs to be expanded before it can truly be said to describe +ABST, +MASS nouns in general.

Accounts of the meaning of items that are lexically +MASS can be divided into two sorts: those that are viewed as part of formal semantics and those that think of give a more "ordinary language" account of the meaning. The former will try either to employ existing constructs from formal semantics or else will recommend the use of some novel logical or mathematical machinery to be incorporated into formal semantics. The latter will offer more of a "dictionary sense" to these lexical entries, often phrased in terms of some set of cross-cultural semantic primitives.

This latter methodology has been most deeply developed within the Natural Semantic Metalanguage (NSM) framework of Wierzbicka (1996). Indeed, many of the works within this framework have concerned themselves directly with the issue of the meaning of +MASS terms, and they have outlined a quite rich structure within the category of +MASS nouns. (See in particular Wierzbicka 1988; Goddard and Wierzbicka 2002; Goddard 2009.) Although there are many who find the NSM framework to be wanting (e.g., Barker 2003; Matthewson 2003; Riemer 2006; Jackendoff 2007), the work has produced the most detailed descriptions of different types of +MASS (and +COUNT) nouns and ought to be studied for that reason alone.

Within the formal semantic group of theories, it has long been recognized that standard first-order logic seems unsuited to representing sentences involving mass terms. For example, even if one grants that (12a) is to be represented as (12b), nonetheless Tarski's classic (13a) can't reasonably be represented as (13b), because there are no plausible values for *x*:

(12) a. Men are mortal.
b. $\forall x(Man(x) \supset Mortal(x))$

(13) a. Snow is white.
b. $\forall x(Snow(x) \supset White(x))$

For, what could be the value of *x* in "For all *x*, if x is snow, then *x* is white"? Intuitively, we want it to be "snowy stuff," but the idiom of classical logic is committed to the values being objects/things/entities—elements of the domain. However, those are what is designated by +COUNT nouns, not +MASS nouns.

Earlier attempts invoked relational constants into the first-order language to accommodate +MASS terms. Parsons (1970), for example, used names for substances and the relations *quantity-of* and *consitituted-by* as well as a "substance-forming operator." Burge (1975) considers two different theories along these lines: a "relational" account that analyzes sentences like (14a) as (14b).

(14) a. This ring is now gold
b. Gold(*r*, now)

He rejects this account because it takes the basic individuals to be stages of objects. His preferred account employs the basic Three-place relation of '*x* constitutes *y* at (time) *t*', analyzing (14a) as

(15) $(\exists x)(\text{Gold}(x) \wedge C(x, r, \text{now}))$

One or the other of these two general approaches have been adopted by a wide range of theorists who want to employ as much of ordinary first-order logic as possible.

An important topic concerns the "dual life" that many (most? all?) nouns lead in English and similar languages. What should the semantic value of the lexical item be for such nouns? Consider *chocolate*, for example. Should its semantic value be the set of chocolates? But of course not only do we have the chocolates that are delivered to one's lover on Valentine's Day but there is the chocolate that they are made from. Here is where a first-choice point arises. Should we take the set of chocolates to be basic and somehow "derive" the chocolate from them? Or should we take the chocolate stuff as basic and derive the set of chocolates from that? Or should there be two separate meanings that are each basic? Or should there be one meaning that is "unspecified for +MASS/+COUNT" thereby implicitly including both meanings?

The literature on mass nouns seems to be committed to taking one of the meanings as basic and deriving the other. The process is called "coercion" and is said to be triggered by features of the surrounding linguistic context (or, sometimes, by the nonlinguistic context). For example, if the mass meaning is taken as primitive, then a sentence like *Abelard gave Heloise seven chocolates* is said to have 'chocolate' coerced into a +COUNT meaning by the presence of the plural and the number modifier. If the sentence were *Each chocolate was made by hand*, then the presence of 'each' will coerce the basic mass meaning into a count meaning. Alternatively, if the count meaning is taken as primitive, then a sentence like *The box contained two kilos of chocolate* would be seen as coercing the basic count meaning into the mass meaning by means of the measure phrase 'two kilos of'.

It is never clear why one meaning vs. the other is taken as primitive. It is as if the theorists have some special insight into what the true, real, and underlying meaning is. It seems unlikely to be due merely to frequency; maybe it is some combination of frequency and saliency. When examples are given, they do not use nouns like *chocolate* but rather *house, child, car, . . .* . It then seems more plausible to say that *Sally owns too much house!* employs a coerced mass meaning of *house*—which is "really" a +COUNT

term. But this seems to become a more suspect attitude when one views the full range of "dual life" terms. The list in (Huddleston and Pullum 2002: p. 335) shows such a truly wide variety of these terms that one would hesitate to employ the notion of 'coercion' so rapidly.

Some theorists have thought that there should be a distinction made within +MASS nouns for those that are homogeneous vs. those that are "atomic." The sort of distinction these theorists have in mind is between words like *blood* and *furniture*. We have already seen in conjunction with (10) above that Huddleston and Pullum (2002) wish to make these form separate categories within −COUNT, on the basis that these "aggregate terms" are true of "very different types of things." This issue of how a mass term can be a cover term for a variety of different subtypes, each of which is a count term, is discussed in the psychology literature under the heading of "mass nouns as superordinate terms" (see Markman 1985; Wisniewski et al. 1996, 2003; Takatori and Schwanenflugel 2008), with an emphasis on whether there is some basis "in reality" for the notion of an 'individual' vs. 'stuff'. This is often called "natural atomicity." We should notice that, while the atomic parts of furniture are rather large—making it clear that the natural atomicity has been fulfilled—the atomic parts of cutlery are smaller and those of gravel are very small indeed. And as I mentioned above, pretty much every English speaker believes that there are atomic parts of *any* purported mass term, even such prototypical ones as 'water' or 'blood'. What *does* seem true, however, is that for some mass nouns, 'gravel' perhaps or maybe 'coffee', just what counts as atomic parts is "vague." The idea is that while there are clear cases of the atomic parts of (most?) +COUNT nouns, and of some +MASS nouns (the clear-cut "aggregates"), and while some words perhaps designate "complete homogeneity" (maybe 'space'?), these endpoints merge into one another in the same way that any of the traditional vague predicates do. And then the issue of what nouns should be +COUNT and which should be +MASS is of the same nature as vagueness: here, the vagueness of what counts as an atomic part. This line of research is pursued by Chierchia (2010).

7 Some Diachronic and Cross-Linguistic Data

Chierchia (2010) gives a very helpful three-way division of how various languages deal with the +MASS/+COUNT distinction. Without insisting on the exhaustivity of its classification, or even on the ultimate "truth" of its vision, we can nonetheless use the labels to give general characterizations. According to this division, the world's languages fall into one of the following three groups with regard to +MASS/+COUNT.

1. Number marking languages, which have overt number features that obligatorily appear on nouns. Here the +MASS/+COUNT distinction applies to the nouns directly. (Most?) Indo-European languages, e.g., current English, are such languages.
2. Classifier languages, which do not have obligatory number marking on nouns (and arguably do not have a singular/plural contrast at all on nouns). Lexical nouns in such languages could be viewed as +mass, although there is a +mass/+count distinction that is active more generally. (For this reason it might be better to view the lexical nouns as unspecified for +mass/+count.) The classifiers in these languages enforce the +mass/+count distinction, but at the level of an entire "classified noun phrase." (Most?) Asian languages, e.g., Mandarin, Japanese, and Korean, are such languages.

3. Languages lacking both obligatory number marking and obligatory classifier systems. Various Amerindian languages, various South American languages, and various Austronesian languages are such languages. It is not clear whether these languages can be said to have a +MASS/+COUNT distinction, although some scholars have argued for it.

Toyota (2009) reports that a close study of the use of mass and count terms in the five recognized phases of the development of the English language shows that there have been substantial changes in the counting system of English, especially in the use of classifiers. Apparently, the earlier English did not make a distinction between mass and count nouns, using classifiers exclusively and therefore having the nouns all be +MASS. The +MASS/+COUNT distinction eventually emerged around around 1500. Toyota concludes that in fact English changed from a classifier language to the current number-marking, non-classifier language, and as part of this change came to mark a +MASS/+COUNT distinction within lexical nouns.

The Chinese languages and the Korean-Japanese languages are often argued as *not* making a +MASS/+COUNT distinction within the lexical noun, because in these languages *no* noun can directly combine with numerals. Instead, a classifier—a word that indicates a way to "individuate" what is being discussed—is always needed. The classifier might designate a measure, or some container, or some shape (etc.) that the referent of the noun is to have. Many writers, e.g., Hansen (1976), Sharvy (1978), Krifka (1995), Chierchia (1998a, b), have concluded that the referent of the noun is therefore to be understood as "mass stuff," waiting to be "classified" into an object or a portion or some shape, etc. This is true for nouns such as 'man' as well as ones like 'water'. And hence, the meaning of all lexical nouns is +MASS. However, the work of Cheng and Sybesma (1999) has convinced many that the appropriate place to look for the +MASS/+COUNT distinction in these languages is the classifier system itself. And thus Chierchia (2010) can now write that although it is *possible* to view the lexical nouns as +MASS, "there is a +MASS/+COUNT distinction that is active more generally."

Other languages have fewer syntactic constructions that would give clues as to whether a noun is +MASS or +COUNT. There are languages that have number marking (e.g., singular/plural), but it is not obligatory to use it. And there are classifier-like languages where the use of the classifiers is not obligatory. More notably, there are languages where there in fact is *no* number marking or classifiers; and languages that have no determiners (like 'a(n)' or 'sm') at all. There are languages where all quantification is adverbial, and not over nouns: One says 'Bears are always dangerous' rather than 'All bears are dangerous'. And there are even languages where one uses the same construction to describe the result of adding two bowls of water together to form one bowl as would be used to consider putting two monkeys in the same vicinity to each other.

An ever-increasing literature on cross-linguistic comparisons of mass/count are available. Good starting points are Massam (2012) and Doetjes (2012).

8 Concluding Remarks

We have seen that the +MASS/+COUNT distinction has been thought to have philosophical ramifications within metaphysics, both as to the ultimate nature of reality and to

issues of identity and reidentification. The extension of the +MASS/+COUNT distinction to the realm of verb phrases also yields a philosophically interesting area of investigation into differentiation among processes, events, achievements, and so on. We have also seen that the existence of mass nouns in natural language has challenged traditional first-order logic as a representation language, and sparked development of mereological and latticetheoretic approaches as representational media.

There are two standard approaches to the +MASS/+COUNT distinction: syntactic and semantic. But there appear to be severe difficulties with making either of them into a general theory. Furthermore, the facts surrounding how +MASS/+COUNT is manifested in languages other than the Indo-European ones might seem to show that the basis and rationale for making the distinction—and perhaps any philosophical consequences that might seem to follow from the distinction—are not really valid as claims about reality in general or about how people might conceptualize reality.

One topic not broached in this survey is the psycholinguistic work done on the +MASS/+COUNT distinction. Most of this work has been done in English, and even that done in other languages has mostly concentrated on numbermarking languages with only a small amount done with classifier languages and none at all done in other languages. Nonetheless, it seems that the results are important for a universal theory of +MASS/+COUNT. Further philosophical studies (and further linguistic studies) of the +MASS/+COUNT distinction in English will need to be cognizant of this work, for it provides a touchstone of empirical reality in the way speakers employ the distinction, and it cannot be ignored in favor of "a more beautiful theory." There is a vast amount of such work, only some of which has been mentioned here. See Barner and Snedecker (2005), Markman (1985), Wisniewski et al (1996, 2003), Takatori and Schwanenflugel (2008) and all the works mentioned in those works.

References

Bach, E. (1986a). The algebra of events, *Linguistics and Philosophy* 9: 5–16.

Bach, E. (1986b). Natural language metaphysics, in R. B. Marcus, G. Dorn, and P. Weingartner (eds.), *Logic, Methodology, and Philosophy of Science, VII*, North Holland, Amsterdam, pp. 573–595.

Barker, C. (2003). Paraphrase is not enough, *Theoretical Linguistics* **29**: 201–209.

Barner, D. and Snedeker, J. (2005). Quantity judgements and individuation: Evidence that mass nouns count, *Cognition* **97**: 41–65.

Bunt, H. (1979). Ensembles and the formal semantic properties of mass terms, in F. J. Pelletier (ed.), *Mass Terms: Some Philosophical Problems*, Reidel, Dordrecht, pp. 249–277.

Bunt, H. (1985). *Mass Terms and Model Theoretic Semantics*, Cambridge UP, Cambridge.

Burge, T. (1972). Truth and mass terms, *Journal of Philosophy* **69**: 263–282.

Burge, T. (1975). Mass terms, count nouns and change, *Synthèse* **31**: 459–478. Reprinted in (Pelletier 1979, pp. 199–218).

Cheng, L. and Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP, *Linguistic Inquiry* **30**: 509–542.

Chierchia, G. (1998a). Plurality of mass nouns and the notion of 'semantic parameter' , in S. Rothstein (ed.), *Events and Grammar*, Kluwer, Dordrecht, pp. 53–103.

Chierchia, G. (1998b). Reference to kinds across languages, *Natural Language Semantics* **6**: 339–405.

Chierchia, G. (2010). Mass nouns, vagueness and semantic variation, *Synthèse* **174**: 99–149.

Doetjies, J. (2012) Count-mass distinctions across languages, in C. Maienborn, K. von Heusinger, and P. Portner (eds.), *Semantics*, de Gruyter, Berlin.

Gillon, B. (1992). Towards a common semantics for English count and mass nouns, *Linguistics and Philosophy* **15**: 597–640.

Gillon, B. (1999). The lexical semantics of English count and mass nouns, in E. Viegas (ed.), *The Breadth and Depth of Semantic Lexicons*, Kluwer, Dordrecht, pp. 19–37.

Goddard, C. (2009). A piece of cheese, a grain of sand: The semantics of mass nouns and unitizers, in F. J. Pelletier (ed.), *Kinds, Things and Stuff*, Oxford UP, New York, pp. 132–165.

Goddard, C. and Wierzbicka, A. (2002). *Meaning and Universal Grammar: Theory and Empirical Findings*, John Benjamins, Philadelphia.

Goddard, C. and Wierzbicka, A. (2010). The semantics of 'abstract nouns': Modern NSM analyses and the thinking of John Locke and Jeremy Bentham. Submitted.

Hansen, C. (1976). Mass nouns and 'a white horse is not a horse', *Philosophy East and West* **26**: 189–209.

Huddleston, R. and Pullum, G. (2002). *The Cambridge Grammar of the English Language*, Cambridge UP, Cambridge.

Jackendoff, R. (2007). Conceptual semantics and natural semantic metalanguage theory have different goals, *Intercultural Pragmatics* **4**: 411–418.

Krifka, M. (1995). Common nouns: A contrastive analysis of English and Chinese, in G. Carlson and F. J. Pelletier (eds.), *The Generic Book*, University of Chicago Press, Chicago, pp. 398–411.

Landman, F. (1991). *Semantic Structures*, Kluwer, Dordrecht.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice theoretical approach, in R. Baüerle, C. Schwarze, and A. von Stechow (eds.), *Meaning, Use and Interpretation*, de Gruyter, Berlin, pp. 303–323.

Markman, E. (1985). Why supordinate category terms can be mass nouns, *Cognition* **19**: 31–53.

Massam, D. (2012). *Count and Mass across Languages*, Oxford UP, Oxford.

Matthewson, L. (2003). Is the meta-language really natural?, *Theoretical Linguistics* **29**: 263–274.

McCawley, J. (1975). Lexicography and the count-mass distinction, *Berkeley Linguistic Society, Vol. 1*, pp. 314–321. Reprinted in J. McCawley (ed.), *Adverbs, Vowels, and Other Objects of Wonder*, University of Chicago Press, Chicago, 1979, pp. 165–173.

Moltmann, F. (1998). Part structures, integrity, and the mass–count distinction, *Synthèse* **116**: 75–111.

Montague, R. (1973). Reply to Moravcsik, in K. Hintikka, J. Moravcsik, and P. Suppes (eds), *Approaches to Natural Language*, Reidel, Dordrecht, pp. 289–294. Reprinted as the proper treatment of mass terms in English in (Pelletier 1979, pp. 173–178).

Mourelatos, A. (1978). Events, processes and states. *Linguistics and Philosophy* **2**: 415–434.

Moravcsik, J. (1973). Mass terms in English, in K. Hintikka, J. Moravcsik and P. Suppes (eds.), *Approaches to Natural Language*, Reidel, Dordrecht, pp. 263–285.

Nicolas, D. (2010). Towards a semantics for mass expressions derived from gradable expressions, Recherches Linguistiques de Vineennes **39**: 163–198.

Ojeda, A. (1993). *Linguistic Individuals*, CSLI Press, Stanford.

Parsons, T. (1970). An analysis of mass and amount terms, *Foundations of Language* **6**: 363–388. Reprinted in (Pelletier 1979, pp. 167–171).

Pelletier, F. J. (1974). On some proposals for the semantics of mass terms, *Journal of Philosophical Logic* **3**: 87–108.

Pelletier, F. J. (1975). Non-singular reference: Some preliminaries, *Philosophia* **5**: 451–465. Reprinted in (Pelletier 1979, pp. 1–14).

Pelletier, F. J. (ed.) (1979). *Mass Terms: Some Philosophical Problems*, Kluwer Academic Pub., Dordrecht.

Pelletier, F. J. and Schubert, L. (1989/2003). Mass expressions, in F. Guenthner and D. Gabbay (eds.), *Handbook of Philosophical Logic, 2nd edition. Vol. 10*, Kluwer, Dordrecht, pp. 249–336. Updated version of the 1989 version in the 1st edition of *Handbook of Philosophical Logic*.

Quine, W. V. (1960). *Word and Object*, MIT Press, Cambridge, MA.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*, Longman, New York.

Riemer, N. (2006). Reductive paraphrase and meaning: A critique of Wierzbickian semantics, *Linguistics and Philosophy* **29**: 347–379.

Sharvy, R. (1978). Maybe English has no count nouns: Notes on Chinese semantics, *Studies in Language* **2**: 345–365.

Takatori, Y. and Schwanenflugel, P. (2008). Superordinate category terms and mass-count noun status, *Linguistic Anthropology* **2**: 199–209.

Toyota, J. (2009). When the mass was counted: English as classifier and nonclassifier language, *Skase Journal of Theoretical Linguistics* **6**: 118–130.

Wierzbicka, A. (1988). Oats and wheat: Mass nouns, iconicity, and human categorization, *in* A. Wierzbicka (ed.), *The Semantics of Grammar*, John Benjamins, Amsterdam, pp. 499–560.

Wierzbicka, A. (1996). *Semantics: Primes and Universals*, Oxford UP, Oxford.

Wisniewski, E., Imai, M. and Casey, L. (1996). On the equivalence of superordinate concepts, *Cognition* **60**: 269–298.

Wisniewski, E., Lamb, C. and Middleton, E. (2003). On the conceptual basis for the count and mass noun distinction, *Language and Cognitive Processes* **18**: 583–624.

3.11
INDEXICALS AND DEMONSTRATIVES

Allyson Mount

Indexicals are as puzzling as they are common. We use them effortlessly all the time, yet upon closer inspection their very nature is the subject of vigorous philosophical debate. Indexicals are expressions that pick out different people, places, or things in different contexts. They include pronouns like 'I', 'she', and 'it'; temporal expressions such as 'now' and 'today'; locational terms like 'here'; and adjectives such as 'actual' and 'present'. Expressions like 'nearby' and 'local' are sometimes considered indexicals as well, though it is contentious whether indexicality should be construed broadly enough to include such expressions. Demonstratives, such as 'this' and 'that', are a subset of indexicals, those typically accompanied by pointing or other gestures to identify the referent. Both indexicals in general and demonstratives in particular give rise to puzzles about linguistic meaning and reference.

Puzzles and Problems

Suppose Nadia and Leon have arranged to meet at a coffee shop at noon. Knowing that she has a tendency to be late, Leon calls Nadia on her cell phone ten minutes before the appointed time and asks whether she is close to the coffee shop yet. Nadia, who is early for once, answers, "I am here now." In this context, 'I' designates Nadia, 'here' designates the coffee shop, and 'now' designates the time of her utterance (ten minutes before noon). Her utterance seems clear, unambiguous, and straightforward.

Now suppose Nadia knows that Leon planned to get an oil change on his way to their meeting. After chatting for a minute, she asks how things went at the garage. Leon replies, "I am still here now. There was a long line." In this context, 'I' designates Leon, 'here' designates the garage, and 'now' designates the time of his utterance (nine minutes before noon). Like Nadia's statement, Leon's reply seems clear, unambiguous, and straightforward.

This exchange is entirely ordinary and commonplace. Yet even this simple conversation raises several important questions about indexical meaning and reference. First, what do the two utterances of 'I', 'here', and 'now' have in common, given that they designate different people, places, and times? How can a philosophical theory capture the invariant meaning of indexicals while still respecting the fact that indexicals do, by their very nature, pick out different things in different circumstances? I will call this the

Puzzle of Invariant Meaning. Unlike some of the other puzzles described below, this one has a solution that is widely (though not unanimously) accepted.

The most obvious candidates for the invariant meaning of 'I', 'here', and 'now' are that 'I' always refers to the speaker or writer of the utterance, 'here' always refers to the place of utterance, and 'now' always refers to the time of utterance. That seems to explain what is going on in the exchange between Nadia and Leon, and in most other ordinary cases. Indeed, it seems like it must be impossible to utter 'I am here now' falsely, or to utter 'I am not here now' truly. But there is at least one situation in which it *is* apparently possible to utter 'I am not here now' truly: on answering machine greetings. If the straightforward rules above accurately capture the invariant meaning of these indexicals, how can we explain this phenomenon? This puzzle is commonly called the Answering Machine Paradox, although similar problems arise with handwritten notes, even when no mechanical technology is involved.

With demonstratives, things become even trickier. Suppose that Susan is at an animal shelter, choosing a cat to adopt. The cats are all tangled together in a playful mass. Susan, who has a strong preference for cats that are a certain shade of gray, says "I want that [pointing to the light gray head of a cat] instead of that [pointing to the darker gray hindquarters of a cat]." Unbeknownst to her, she has pointed at different parts of the same cat, instead of at separate cats. How can we explain the fact that her two utterances of 'that' seem to differ in significance, even though her gestures pick out the same cat? This is the Problem of Multiple Utterances for demonstratives, and it has several variations.

Other puzzles arise when there is a discrepancy between the speaker's intentions and the object her gesture actually identifies, and when complex demonstrative phrases are used instead of bare demonstratives. If the gray cat runs away and an orange cat slips into its place just as Susan says "I want that cat," does 'that cat' refer to the gray cat as she intended, or to the orange cat she pointed at? What if she inadvertently points to a ferret instead of a cat? The Problem of Mistaken Gestures involves some kind of mismatch between the gesture and the speaker's intentions. The Puzzle of Quantificational Uses arises when attempting to explain both cases where complex demonstratives are apparently directly referential (as in the previous example) and cases where they seem to function as quantifiers.

This is certainly not an exhaustive list of puzzles involving indexicals and demonstratives. Notably, puzzles that are primarily about the nature of *beliefs* with indexical content are omitted, though beliefs and their linguistic expression are often intertwined. (See Lewis 1979 and Perry 1979.) However, the puzzles above do highlight some of the main challenges that philosophers of language face in providing a compelling account of indexicality. The main strains of thought in response to these puzzles will be traced in what follows.

Kaplan's Theory

On one hand, all utterances of 'I' (or 'today' or 'here', etc.) clearly have some aspect of meaning in common. On the other hand, different utterances of these expressions are used to refer to different things. The Puzzle of Invariant Meaning raises fundamental questions about the nature of indexicals precisely because these two facts are indisputable yet difficult to reconcile. While meaning and reference are different phenomena, with nonindexical terms meaning generally affects reference in fairly straightforward

ways. The relationship between the meaning and reference of indexicals is more complex, but sorting it out is a necessary first step in any theory of indexicality.

A great deal of headway has been made on this problem. Recognition that there *is* a problem to sort out goes back at least to the early 20th century. In "The Thought" (1918), Gottlob Frege comments that "If someone wants to say the same today as he expressed yesterday using the word 'today', he must replace this word with 'yesterday'. Although the thought is the same, its verbal expression must be different so that the sense, which would otherwise be affected by the differing times of utterance, is readjusted" (Frege 1918: 40). Subsequent scholarship has questioned whether some of Frege's claims about the expressions later known as indexicals are consistent with his overall theory of language (Perry 1977). Whether or not they are, Frege's explicit recognition that such expressions are more complex than they first appear certainly contributed to a rising awareness of the issue.

The Puzzle of Invariant Meaning calls attention to the fact that a multi-faceted approach is needed to capture how indexicals work. Like many other linguistic puzzles, this one was initially addressed by logicians who sought to represent language systematically in order to clarify what contribution different expressions make to truth conditions. Building on the work of Rudolf Carnap and others, Richard Montague (1974) developed the idea of using an index to represent linguistic context. At the time of his writing, much philosophical energy was being devoted to possible worlds semantics, where a formal distinction was made between the world in which an utterance is actually made and other possible worlds in which that utterance could be evaluated. The way context was represented varied from theory to theory, but the essential idea for present purposes is that context could be formally represented as an index that contained places for times, places, people, worlds, and whatever other factors were needed to correspond to aspects of the natural language in question. Since English contains tense operators, the index needs a place for time; since it contains modal operators, the index needs a place for world; since it contains the word 'you', the index needs a place for addressee, and so on.

Montague's work paved the way for David Kaplan to expand and refine the idea of using an index to represent linguistic context. In his highly influential paper "Demonstratives" (1989a, though first circulated in 1977), Kaplan develops a system of double-indexing to capture both the invariant meaning of indexicals and the way in which their reference shifts from context to context. The importance of double-indexing for 'now' was recognized earlier by Hans Kamp (1971). However, a full appreciation that Kamp's insights applied to indexicals more generally did not immediately follow, and it took Kaplan's wide-ranging and systematic development to clarify and broaden the idea.

Kaplan conceives of context as an n-tuple with a place for agent, time, location, and world. He distinguishes between contexts of use and circumstances of evaluation. Contexts of use are occasions on which an utterance is made. The referent of "pure" indexicals depends on the context of use, he claims, and the referent of demonstratives depends on the associated demonstration (Kaplan 1989a: 492). Once the referent has been fixed, however, the proposition can then be evaluated with respect to other possible circumstances of evaluation. Kaplan emphasizes that demonstratives are directly referential. Once the referent of an indexical is established in the actual context of use, it is the very same individual or object that is relevant to the evaluation of the proposition in other circumstances of evaluation, rather than the individual the indexical would have picked out if the utterance had been made in those other circumstances. As we shall see,

this feature of Kaplan's account has recently been challenged on various grounds, most notably by Jeffrey King's quantificational theory of complex demonstratives.

The clarification that context of use differs from circumstances of evaluation led Kaplan to distinguish between character and content. The *character* of an indexical is a rule for determining the content in each context of use. This can be represented by a function from contexts to contents. In essence, the character captures the meaning that remains constant. *Content*, on the other hand, is "what is said" (the proposition, for assertions) by the utterance. It varies from context to context and can be represented by a function from circumstances to extensions.

The character rule for 'I', for instance, states that 'I' is a directly referential expression that refers, in each of its utterances, to the person who utters it (Kaplan 1989a: 520). In a particular context of use, the content of 'I' is the individual determined by the character rule—i.e., the speaker or writer. What distinguishes indexicals from non-indexical expressions is that indexicals have a context-sensitive character instead of a fixed character. Since character is determined by linguistic convention and remains constant from use to use, Kaplan suggests that character is more aptly described as the meaning of indexicals.

This brings us to a potential solution to the Puzzle of Invariant Meaning: Character rules capture the aspect of meaning that remains the same from use to use, and Kaplan's theory explains the relationship between the meaning and referents of indexicals. While Kaplan's focus is on providing a theoretical framework for indexicals and demonstratives rather than on specifying what the character rules are for each indexical, one could work within his framework to do so.

The basic tenets of Kaplan's theory have been widely accepted, but it is not without its detractors. Most critiques are not a wholesale rejection of the view but disputes about various details. For instance, a division between "pure" indexicals and true demonstratives is central for Kaplan. Pure indexicals always refer to the relevant feature of the context, as represented by corresponding part of the n-tuple. For these indexicals, Kaplan says, "[N]o associated demonstration is required, and any demonstration supplied is either for emphasis or is irrelevant . . . The linguistic rules which govern their use fully determine the referent for each context" (Kaplan 1989a: 491). True demonstratives, in contrast, require an associated demonstration, which Kaplan says is "typically, though not invariably, a (visual) presentation of a local object discriminated by a pointing" (Kaplan 1989a: 490). Other theorists have discussed Kaplan's contrast between pure indexicals and true demonstratives in terms of "automatic" versus "discretionary" indexicals. Some version of this distinction has been accepted by many philosophers of language, though the details vary and the distinction is not uncontentious (for discussion, see Wettstein 1984; Recanati 2001; Perry 2001; Bach 2005; and Mount 2008).

This brings us to the next puzzle: the Answering Machine Paradox. It is a challenge for proponents of the pure indexical view because they must either insist that utterances of 'I am not here now' are always false despite a common intuition that it can be uttered truly on answering machine recordings, or they must explain why some feature of the answering machine scenario justifies a different view of indexical reference for this particular type of case.

Kaplan explicitly states that 'I am here now' cannot be uttered falsely (Kaplan 1989a: 509), and he appears equally committed to the claim that 'I am not here now' cannot be uttered truly. In a footnote, he acknowledges that recorded messages may introduce additional difficulties, but he does not see this as a challenge to his distinction between

pure indexicals and demonstratives (Kaplan 1989a: 491 fn 12). Others have seen it as a real puzzle in need of resolution, if not necessarily a true paradox. Alan Sidelle (1991) suggests that answering machine recordings should be treated as deferred utterances, where the utterance actually occurs at the time the message is played, rather than at the time it is recorded. This allows him to say that 'now' does refer to the time of utterance (preserving Kaplan's character rule), yet it explains how 'I am not here now' can be uttered truly, since 'now' refers to the time of playback.

Stefano Predelli (1998) proposes another alternative: utterances involving 'I', 'here', and 'now' should be interpreted with respect to the context that is intended by the speaker, rather than with respect to the context of utterance. In many cases the intended context of interpretation will just be the context of utterance, but the two do not necessarily coincide. Corazza et al. (2002) explore the consequences of Predelli's view, arguing that it is implausible because it allows 'I' to refer to virtually anyone, simply because the speaker intends so. Instead, they propose that the relevant context is the one determined by convention. Since answering machines are devices conventionally understood to broadcast messages to be interpreted at the time of playback, it is the time of playback that is relevant—not simply because of the speaker's intention, but because social conventions dictate such an interpretation.

Others have joined in the debate, responding to the Answering Machine Paradox either by modifying Kaplan's character rules to account for recorded messages or by replacing some other element of his theory. Regardless of which view one favors, some response to the puzzle is needed.

Simple Demonstratives

The criteria for what makes something a demonstration, and what happens when the demonstration is ambiguous or various other contingencies arise, have been a matter of continuing disagreement. The Problem of Mistaken Gestures arises when the speaker uses a gesture to aid hearers in identifying the intended referent, but for some reason things go astray. This could happen when the intended demonstratum suddenly moves, or when the speaker hallucinates and there is no object in the indicated spot at all, or when the gesture is thwarted in another way. One strain of thought is that speaker intentions are private and internal, so intentions cannot play a constitutive role in what a particular utterance of a demonstrative refers to, though they certainly are important in what the speaker means to communicate. This type of view usually gives more weight to publicly observable gestures whenever gestures and intentions conflict. Another strain of thought is that speaker intentions—or at least a certain sort of intention—must figure prominently in any theory of demonstratives. This type of view leads to questions about whether intentions are relevant to the semantics of demonstratives or just to pragmatics (a topic taken up in a later section). Kaplan addresses some concerns about the role of demonstrations and intentions in "Dthat" (1978a) and "Afterthoughts" (1989b). More recently, Kent Bach and Marga Reimer rekindled interest in the topic, discussing various cases where speaker intentions and gestures diverge (Bach 1987 and 1992; Reimer 1991a and 1991b).

Like the Problem of Mistaken Gestures, the Problem of Multiples Utterances comes in several varieties. One version involves co-reference, as when Susan points and utters 'that' twice, unaware that her gesture picks out different parts of the same cat. A similar and more widely discussed problem of co-reference arises for proper names and is known

as Frege's Puzzle. The solution Frege (1892) advocates for proper names is to distinguish between *Sinn* and *Bedeutung*, or sense and reference. Proper names can have different senses even when they refer to the same person, he argues, and this explains how someone could have differing attitudes about a proposition depending on how it is expressed. One might try to explain the related version of the Problem of Multiple Utterances similarly. When Susan unknowingly points to different parts of the same cat and says "I want that [pointing to the light gray head of a cat] instead of that [pointing to the darker gray hindquarters of a cat]," the cognitive value of each utterance of 'that' (paired with a demonstration) differs, even though they refer to the same object. However, it is far from clear that this strategy alone can solve the problem.

A second version of the Problem of Multiple Utterances is highlighted by David Braun (1996). Braun argues that Kaplan's theory of demonstratives cannot adequately account for utterances in which the same demonstrative is used multiple times to refer to *different* objects. The trouble arises because Kaplan identifies the linguistic meaning of demonstratives with their character, and character is a function from contexts to contents. But if character is the linguistic meaning that remains constant from use to use, it seems that any two occurrences of the same demonstrative in the same context should have the same content (refer to the same object). Yet this flies in the face of the observation that sentences like "That is bigger than that" or "That [pointing to an aged book] is older than that [pointing to an antique vase]" can be uttered truly even though nothing is bigger or older than itself. While Kaplan does propose a formal language that treats different occurrences of 'that' as associated with different demonstrations, thus having different referents, Braun argues that Kaplan's formal treatment avoids the problem but implies that character alone does not capture the linguistic meaning of demonstratives.

Braun considers and rejects a modification of Kaplan's theory that allows context shifts in midsentence. This modification seeks to solve the problem by saying that the context in which the first 'that' is uttered is relevantly different from the context in which the second 'that' is uttered. In other words, the context shifts within the short time it takes to utter the sentence, allowing each occurrence of the demonstrative to refer a different object while still maintaining that 'that' has a single character that is its linguistic meaning. Braun rejects this alternative and instead argues that demonstratives have three distinct types of meaning, not two. First, there is content, as in Kaplan's theory. Second, there is linguistic meaning that remains constant from context to context. Third, there is character that a demonstrative acquires when associated with a demonstration. The linguistic meaning of the expression, when combined with a demonstration, yields a character. That character then works with the context to yield a certain content or referent. This avoids Kaplan's problem by allowing different demonstrations accompanying each utterance of 'that' to affect its character. So while the linguistic meaning of 'that' remains constant, character (and content) differs.

The Problem of Mistaken Gestures, the Problem of Multiple Utterances, and critiques such as Braun's all make it clear that even simple demonstratives raise challenging questions. Not surprisingly, complex demonstratives—noun phrases like 'that cat' and 'this green crayon'—raise additional questions about how demonstratives work.

Complex Demonstratives

A version of the Problem of Mistaken Gestures also arises for complex demonstratives. This happens when the object demonstrated is not of the sort described by the phrase, as

when Susan utters 'that cat' while pointing at a ferret. In one sense, this version of the problem is clearer because the use of a noun adds linguistic constraints, making it extremely unappealing to think that the speaker's intentions alone could ever carry more weight than the observable features of the utterance. But things are complicated by the fact that the speaker's intentions and gestures are even more misaligned than with simple demonstratives, raising additional questions about reference failure. Providing a plausible theory of what is going on in cases of this sort is especially difficult because so many variations exist.

The Problem of Mistaken Gestures is certainly challenging, but the discussion is generally framed within the familiar Kaplan-style framework, though there may be considerable disagreement about the details. The Puzzle of Quantificational Uses is based on a more sweeping critique which, if substantiated, would undermine Kaplan's whole approach.

It is essential to Kaplan's view that demonstratives (and indexicals in general) are directly referential. Failure to recognize this, Kaplan believes, is a primary reason why previous theories were misguided. King (2001) provides data that is difficult to reconcile with the direct reference view. He argues that the contributions that complex 'that' phrases make to propositions are akin to those made by quantificational noun phrases like 'every skier' and 'most swimmers'. This contrasts with the direct reference approach, which treats such phrases as contributing individuals as the content. Hence, the core of the dispute is whether complex 'that' phrases are referring expressions or quantifiers.

King finds direct reference accounts problematic in many respects. He argues that examples involving what he calls NDNS uses (*no demonstration no speaker reference*), *quantification in*, and *narrow scope* readings all favor a quantificational view. Quantification in occurs in "Every father dreads that moment when his oldest child leaves home." The complex demonstrative phrase is bound by the quantifier 'Every father' and does not refer to any particular individual. Sentences such as "That professor who brought in the biggest grant in each division will be honored" are ambiguous but can be continued in a way that makes it clear that a narrow scope reading is intended (King 2001: 10). Since no particular individual is identified by the complex demonstrative phrase in either of these types of examples, direct reference accounts seem ill-equipped to explain them.

NDNS uses require a bit more explanation. King notes that 'that' phrases can be used without an accompanying demonstration, to talk about something that is not in the speaker's immediate environment, when the speaker has no particular individual in mind. This is an NDNS use. He consider the sentence "That student who scored one hundred on the exam is a genius," uttered by Greg in a situation where Greg has no idea who scored one hundred, but has been assured by his trustworthy teaching assistant that exactly one student received a perfect score. While a direct reference theorist could simply say 'that student who scored one hundred on the exam' contributes a particular object to the proposition, such a response is less plausible when the above sentence is embedded in a belief report. If someone else reports, "Greg believes that that student who scored one hundred on the exam is a genius," the reporter's use of the 'that' phrase is itself an NDNS use. Her claim is true, and King argues that the direct reference theorist lacks the resources to explain its truth. If the complex demonstrative phrase contributes an individual to the singular proposition, then the truth of the reporter's claim would be the result of Greg standing in the belief relation to that proposition. But Greg had only general grounds for his assertion, so King argues that this analysis is strained.

On the whole, King's quantificational view does not entail that speaker intentions are totally irrelevant to the semantics of 'that' phrases. Intentions are relevant, he says,

in the sense that they can determine properties that further restrict the quantification, beyond the restriction by the descriptive material in the phrase itself (King 2001: 34). He also acknowledges that some uses of 'that' phrases are rigid, while others are not. The difference rests on whether it is the speaker's *perceptual intentions* or *descriptive intentions* that determine the property further restricting the quantification. However, he argues against ambiguity approaches, which treat rigid uses as directly referential while accepting a quantificational approach only for nonrigid uses. King claims that a unified quantificational approach is preferable, since his account does give a principled explanation of why some uses are rigid and other are not.

King's view has spawned extensive discussion in recent years, leading to various modifications and clarifications by proponents of both direct reference and quantificational theories. (See, for example, Braun 2008a and 2008b; King 2008a and 2008b; see also Lepore and Ludwig 2000.) The Puzzle of Quantificational Uses is a puzzle of how to reconcile examples that apparently support a direct reference view with examples that apparently support a quantificational view, in a theoretically satisfying way. The debate remains alive and well.

Indexicals and the Semantics-Pragmatics Distinction

No discussion of indexicals and demonstratives would be complete without mentioning the debate over the semantics-pragmatics divide. The debate has implications for many aspects of language, not just indexicals, and the issue is far too complex to do more than just scratch the surface here. (See Chapter 1.2 on Semantics and Pragmatics in this volume.) In brief, the question is whether indexicals themselves are referring expressions in the semantic sense, or whether they are just *used* by speakers to refer in a pragmatic sense.

In the mid-20th century, ordinary language philosophers saw reference as a property of uses of expressions rather than of expressions themselves. Kaplan's theory undermines this ordinary language view, since it implies that it does make sense to speak of indexicals themselves referring—just relative to a context, not simpliciter. He shows that it is possible to give a logic of demonstratives (1978b). We have already seen some of the challenges to Kaplan's theory raised by Braun, King, and others. Behind the scenes in all of this lurk questions about what role context plays. Does it simply provide parameters that can be incorporated into meaning rules of indexicals directly? Should we conceive of context as including demonstrations, and if so, what is the best way to represent them? Are speaker intentions part of context in the semantic sense, or are things like beliefs and intentions extra-linguistic factors that are certainly relevant to communication but not to semantics? Should we abandon talk of indexicals referring in favor of saying that indexical reference is just speaker's reference?

John Perry (2001) provides a theory of indexicals and proper names that addresses some of these foundational issues. He compares referentialist and descriptivist views of indexicals, ultimately drawing insights from both to support his own referential-reflexive view.

The referentialist view of indexicals holds that indexicals designate objects, and statements containing indexicals express propositions about those objects (Perry 2001: 7). A statement containing 'I', for instance, expresses the same proposition as a similar statement where 'I' is replaced with the name of the speaker, since both designate the same person. This view seems to get counterfactual truth conditions right; in evaluating

whether the proposition is true at other possible worlds where the descriptions associated with the person vary, it still seems like *that person* (the speaker who uttered 'I' in the actual world) is the one relevant to evaluation of the proposition, even if she is not speaking in the other world. However, referentialists have difficulty explaining how co-referring terms can differ in cognitive significance, as Frege recognized long ago. If the listener does not know he is speaking with Michelle Obama, he might believe the proposition expressed by "Michelle Obama is Barak Obama's wife" without believing the proposition expressed by "I am Barak Obama's wife" [uttered by Michelle Obama].

Descriptivist views of indexicals come in many variations, but the basic idea is that "statements involving indexicals express propositions that incorporate identifying conditions . . . that derive from the rules of the language" (Perry 2001: 9). Hans Reichenbach's theory of token-reflexives took this approach (Reichenbach 1947). Reichenbach suggests that what indexicals contribute to a proposition is a reflexive content about that token of the indexical itself. For instance, "I prefer tangerines to oranges" really means something like "*The speaker of this token of 'I'* prefers tangerines to oranges." Supporting this view is the intuition that competent speakers of English understand the reflexive content of 'I' even when they do not know who the speaker is. The same approach can be used for other indexicals and demonstratives, with 'here' meaning *the place where this token of 'here' is spoken*, and so on.

Building on earlier work by Arthur Burks (1949), Perry takes insights from both referentialist and descriptivist views and develops what he calls the referential-reflexive theory. According to Perry, both reflexive contents and referential contents are associated with indexicals. Reflexive content explains what is going on in co-reference cases. "I am Barak Obama's wife" and "Michelle Obama is Barak Obama's wife" [uttered by Michelle Obama] differ in their reflexive contents even though 'I' and 'Michelle Obama' refer to the same person. Thus, a token-reflexive view of indexicals seems promising. However, as Perry acknowledges, it is odd to think that when we use indexicals we are making statements about the words themselves. In most ordinary cases, we aren't—we are talking about the actual subject matter, not about words. That is where referential content comes in, contributing to what is officially said by the utterance (Perry 2001: 12–13).

Perry distinguishes between automatic and discretionary indexicals in a now-familiar way, and also between narrow and wide context. He says, "*narrow context* consists of facts about which things occupy the essential roles involved in the utterance, which I will take to be the agent, time, and position. These roles are filled with every utterance . . . The *wide context* consists of those facts, plus anything else that might be relevant, according to the workings of a particular indexical" (Perry 2001: 59). According to Perry, some indexicals (such as 'I') rely only on narrow context, while others rely on wide context. This type of distinction between narrow and wide context has been used by others as well, often in conjunction with a distinction between pure/automatic indexicals and discretionary indexicals/true demonstratives. It allows one to say that narrow context is the type of context relevant to semantics, while still acknowledging the common practice of using the word 'context' in a broader sense to include speaker intentions, beliefs, and other factors.

Space does not permit a full discussion of Perry's theory here. (See O'Rourke and Washington 2007 for commentary.) The key point for present purposes is that the issue of how to conceive of context is often tied to the issue of how to draw the line between semantics and pragmatics, and this in turn affects one's view of indexicals. Bach (2007) has questioned Perry's conception, focusing on two of Perry's assumptions: that utterances

(rather than sentences) have semantic contents, and that indexicals and demonstratives are inherently referential. Pointing out apparent inconsistencies in Perry's view, he charges Perry with blurring the line between semantic "official content" and speaker meaning. Perry accepts Kaplan's view that speakers' directing intentions play a part in the reference of discretionary indexicals. But if that is true, Bach argues, then intentions do so pragmatically. Kaplan's view only works for automatic indexicals, which depend only on the contextual parameters included in his n-tuple. Speaker intentions are *not* part of context. Bach concludes that "the meanings of discretionary indexicals are not utterance-reflexive" (Bach 2007: 408), since discretionary intentions are communicative intentions, and communicative intentions only affect pragmatics, not semantics.

The issue of where to draw the line between semantics and pragmatics is directly related to the Problem of Mistaken Gestures as well. Appealing to a distinction between narrow and wide context solves some problems while exacerbating others. If neither gestures nor intentions are included in narrow context, that may make it more difficult to explain the semantic role of demonstratives. If gestures but not intentions are included in narrow context, that paves the way for treating gestures as taking precedence when they conflict with intentions in simple demonstrative utterances. However, it further confounds attempts to clarify the relationship between complex demonstratives and gestures, since a speaker could use a complex demonstrative phrase while pointing at an object that is not of the sort indicated (for example, pointing to a ferret while uttering 'that cat').

Conclusion

The puzzles we began with remain unresolved, but over the last few decades progress has been made on many fronts. Kaplan-style character rules are leading candidates to address the Puzzle of Invariant Meaning, although modifications may be needed to explain the Answering Machine Paradox and other unusual cases. Quantificational accounts of complex demonstratives provide another sort of explanation altogether. The jury is still out on which approach better accounts for the data. However, at least it is now clear that the examples involved in the Puzzle of Quantificational Uses require explanation, in addition to the more familiar cases where a direct reference account seems promising. This insight sharpens the focus on exactly what remains to be explained.

The Problem of Multiple Utterances highlights the fact that the relationship between the meaning and reference of indexicals remains unclear. How this debate is eventually resolved has implications for other areas of philosophy as well, especially for philosophy of mind and epistemology. Likewise, the Problem of Mistaken Gestures raises difficult questions about the role of gestures and speaker intentions. Aside from being interesting in its own right, this issue is directly related to the ongoing debate about how best to understand linguistic context, and where to draw the line between semantics and pragmatics. With all of these puzzles remaining, it is no surprise that philosophers of language continue to explore many facets of indexicals and demonstratives.

Related Topics

1.1 Extensions, Intensions, Characters, and Beyond
1.2 Semantics and Pragmatics

1.10 Context Sensitivity
2.1 Reference
5.9 Montague Grammar.

References

Bach, K. (1987) *Thought and Reference*, New York: Oxford University Press.
——(1992) "Intentions and Demonstrations." *Analysis*, 52(3): 140–146.
——(2005) "Context ex machine," in Zoltán Gendler Szabó (ed.), *Semantics versus Pragmatics*, New York: Oxford University Press.
——(2007) "Reflections on *Reference and Reflexivity*," in M. O'Rourke and C. Washington (eds.), *Situating Semantics*, Cambridge, MA: MIT Press.
Braun, D. (1996) "Demonstratives and Their Linguistic Meanings." *Noûs*, 30: 145–173.
——(2008a) "Problems for a Quantificational Theory of Complex Demonstratives." *Philosophical Studies*, 140 (3): 335–358.
——(2008b) "Persisting Problems for a Quantificational Theory of Complex Demonstratives." *Philosophical Studies*, 141 (2): 243–262.
Burks, A. (1949) "Icon, Index, and Symbol." *Philosophical and Phenomenological Research*, IX: 673–689.
Corazza, E., Fish, W. and Gorvett, J. (2002) "Who is I." *Philosophical Studies*, 107 (1): 1–21.
Frege, G. 1892. "On Sense and Nominatum," in A.P. Martinich (ed.), *The Philosophy of Language*, 5th edn, New York: Oxford University Press.
——(1918) "The Thought," in A.P. Martinich (ed.), *The Philosophy of Language*, 5th edn, New York: Oxford University Press.
Kamp, H. (1971) "Formal Properties of 'Now'." *Theoria*, 37 (3): 227–273.
Kaplan, D. (1978a) "Dthat," in A.P. Martinich (ed.), *The Philosophy of Language*, 5th edn, New York: Oxford University Press.
——(1978b) "On the Logic of Demonstratives," in A.P. Martinich (ed.), *The Philosophy of Language*, 5th edn, New York: Oxford University Press.
——(1989a) "Demonstratives," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press.
——(1989b) "Afterthoughts," in Matthew Davidson (ed.), *On Sense and Direct Reference: Readings in the Philosophy of Language*, Boston: McGraw Hill.
King, J.C. (2001) *Complex Demonstratives: A Quantificational Account*, Cambridge, MA: MIT Press.
——(2008a) "Complex Demonstratives, QI Uses, and Direct Reference." *Philosophical Review*, 117 (1): 99–117.
——(2008b) "Complex Demonstratives as Quantifiers: Objections and Replies." *Philosophical Studies*, 141 (2): 209–242.
Lepore, E. and Ludwig, K. (2000) "The Sementics and Pragmatics of Complex Demonstratives." *Mind*, 109 (434): 199–240.
Lewis, D. (1979) "Attitudes De Dicto and De Se." *Philosophical Review*, 88 (4): 513–543.
Montague, R. (1974) *Formal Philosophy*, New Haven, CT: Yale University Press.
Mount, A. (2008) "The Impurity of 'Pure' Indexicals." *Philosophical Studies*, 138 (2): 193–209.
O'Rourke, M. and Washington, C. (eds.) (2007) *Situating Semantics: Essays on the Philosophy of John Perry*, Cambridge, MA: MIT Press.
Perry, J. (1977) "Frege on Demonstratives." *Philosophical Review*, 86 (4): 474–497.
——(1979) "The Problem of the Essential Indexical." *Noûs*, 13: 3–20.
——(2001) *Reference and Reflexivity*, Stanford: CSLI Publications.
Predelli, S. (1998) "I Am Not Here Now." *Analysis*, 58(2): 107–115.
Recanati, F. (2001) "Are 'Here' and 'Now' Indexicals?" *Texte*, 127: 115–127.
Reimer, M. (1991a) "Do Demonstrations Have Semantic Significance?" *Analysis*, 51(4): 177–183.
——(1991b) "Demonstratives, Demonstrations, and Demonstrata." *Philosophical Studies*, 63: 187–202.
Reichenbach, H. (1947) *Elements of Symbolic Logic*, New York: The Free Press.
Sidelle, A. (1991) "The Answering Machine Paradox." *Canadian Journal of Philosophy*, 21: 525–539.
Wettstein, H. (1984) "How to Bridge the Gap Between Meaning and Reference." *Synthese*, 58: 63–84.

3.12
INDICATIVE CONDITIONALS

Anthony S. Gillies

1 The Target

Sometimes philosophy gets started by pointing. Setting out the target for theories of *indicative conditionals* is a case in point. So some examples to point at:

(1) a. If the gardener didn't do it, (then) it was the butler.
    b. If your blue marble is in the box, then your red one might be under the couch.

Our target is the class of ordinary *if . . . then . . .* constructions like these. They express conditional information: information about what is or might or must be, if such-and-such is or turns out to be the case.

Indicatives (the name is not great but is entrenched) stand apart from two other sorts of conditional constructions. First, from subjunctive or counter factual conditionals. An Adams-pair:

(2) a. If Oswald didn't kill Kennedy, then someone else did.
    b. If Oswald hadn't killed Kennedy, someone else would have.

The conditional information these traffic in is different: (2a) is an indicative and true; it's counterfactual cousin (2b) is false. Subjunctive conditionals—that is, those with a distinctive tense/aspect marking like (2b)—say something about would or might *have been* if such-and-such *had been*.

(The "subjunctive" marking doesn't go exactly with counterfactuality. Indeed, it's neither necessary (as Anderson (1951)-type examples show) nor sufficient (as sportscasterese-type conditionals show). See von Fintel 1998 (and the references therein) for the status of the connection between "subjunctive" marking and counterfactuality.)

Second, indicatives stand apart from *biscuit conditionals*. Austin's (1956) example:

(3) There are biscuits on the sideboard if you want them.

A fine thing to say, but not normally a thing that expresses conditional informatinon. While it is open to connect a story about indicatives and counterfactuals, and while it is open to connect a story about indicatives and biscuit conditionals, we will be setting such outreach aspirations aside. Having pointed a few times, let's get going.

2 Landscape

Saying that indicatives carry conditional information isn't yet to say very much. We a want a better grip on:

i. Just what that conditional information is. This is a story about what indicatives in natural language mean (what semantic values they have) and how they are used in well-run conversation (what their pragmatic profile looks like).
ii. Just how sentences of natural language express those meanings. This is a story about how *if* interacts with hte rest of out language (that holds good both for compositionally deriving how *if*s express conditional information and for seeing how conditional sentences contribute their conditional meaning to embedding environments).

Suppose that semantic values (whatever they are) determine truth values at points of evaluation (worlds,situations, whatever). This holds, let's assume, for borth iffy sentences. Not wholly innocent—as we'll see—but a place to start.

Let's distinguish between (i) conditional sentences of natural laguage and (ii) conditional connectives of some formal languages that serves to represent the logical forms of conditional sentences. The aim is then to associate an *if*-in-English (via well-behaved mapping that we won't bother with) with an *if*-in-the formal-language that then is associated (via a well-behaved mapping we will bother with) with its semantic value. (There are ways to express conditional information in natural language without resorting to *if*s (in some languages, it's the only way). The issues in this particular ballpark won't matter too much for our purposes so we can (pretty) safely focus on the *if* . . . *then* of English.) For simplicity we'll stick with a simple propositional language (with sentential connectives $\neg$, $\wedge$, $\vee$, $\supset$) together with a connective $(\mathit{if}\ \cdot)(\cdot)$ for the indicative. (Whether indicatives can be represented by a binary conditional connective in a regimented intermediate language is also, as we'll see, uo for grabs.) This indirect route to assigning meanings to indicatives isn't required, is in principle dispensable (assuming our mappings are well-behaved), but makes for an unobstructed view.

One way to make sense out of the idea of conditional information is to let some version of the Ramsey Test guide us. Here's what Ramsey (1929/1990: 155) says: "If two people are arguing 'If p, will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q. . . ." The adding this involves can't be feigned or hypothetical belief. (Thomason is credited with this observation in van Fraassen 1980.) If it were, the Ramsey Test would go wrong:

(4) If my students are cheating in class, then I will not discover it (because they're so clever).

The state I get into by feigning belief in the antecedent won't be one in which I believe the consequent. The right way to understand argumenting for the Ramsey test is restricting a body of information by the information carried by the antecedent. A more neutral version of the Ramsey Test Skirts this: the conditional information carried by (if P) (Q) is true/accurate/acceptable in a situation iff Q is true/accurate/acceptable in that situatuion-plus-the-information-that-P.

Theories of indicatives are constrained by the patterns of intuitive entailment they participate in. So what-intuitively-entails-what is important. We want to explain those patterns as best we can. Consider whether indicatives generally go in for a *deduction theorem* (a.k.a. *conditional proof*):

(5) $\Gamma \supset \{P\} \models Q$ iff $\Gamma \models (\textit{if}\ P)(Q)$

If so, then we have an especially tight connection between what $(\textit{if}\ \cdot)(\cdot)$ means and what we say about entailment. the point is that 'entailment' is as much part of the theoretical machinery as is anything. Take it as given that what we may say about *if*'s can interact in non-trivial ways with what we say about entailment. Then the best route to the best theory may be one in which what we say about entailment bends to the will of the *if*'s as much as the other way around.

A concrete example: what Stalnaker (1975) calls the *direct argument*.

(6) a. Either the butler did it or the gardener did it.
$P \vee Q$
b. So: if the butler didn't do it, then the gardener did.
$(\textit{if}\ \neg P)(Q)$

Seems like (6a) entails (6b). The trouble is that treating such *or*-to-*if* arguments as entailments seems to get us quickly to the conclusion that indicatives mean just what their corresponding material conditionals do. Suppose semantic values are something that determine truth values and that entailment is preservation of truth. And assume the direct argument is an entailment (that $P \vee Q \models (\textit{if}\ \neg P)(Q)$). Then $P \supset Q$ entails the indicative $(\textit{if}\ P)(Q)$.

Argument. Suppose $P \supset Q$ is true (at a world w). Then so is $\neg P \vee Q$. Thus by the direct argument so is $(\textit{if}\ P)\ (Q)$ (at w). (The entailment in the other direction isn't usually disputed.) □

We can follow where the argument leads or we can look for places to get off the boat.

Another example: a *boundedness* argument (Gibbard 1981; Veltman 1985). Indicatives seem to fall somewhere on the spectrum of logical strength between strict implication and material implication: when P entails Q the indicative $(\textit{if}\ P)(Q)$ must be true and the falsity of $P \supset Q$ entails the falsity of the indicative $(\textit{if}\ P)(Q)$. That last bit—that indicatives are bounded from below by material conditionals—is equivalent to saying that they go in for modus ponens. And then there is *import/export*:

(7) a. If the gardener is away and the driver is away, then the mansion is empty.
$(\textit{if}\ (P \wedge Q))(R)$
b. If the gardener is away, then if the driver is away then the mansion is empty.
$(\textit{if}\ P)\ ((\textit{if}\ Q)(R))$

Pairs like these seem to be mutual entailers. Again it seems that indicative conditionals and material conditionals say the same thing, that the truth of $P \supset Q$ is sufficient for the truth of the indicative $(\textit{if}\ P)(Q)$.

Argument. Suppose $\neg P$ is true. Since $(\neg P \wedge P)$ entails Q and indicatives are true when their antecedents entail their consequents, we have that

(8) $(\textit{if}\ (\neg P \wedge P))(Q)$

is true. By import/export

(9) $(if \neg P)((if\ P)(Q))$

is true. And since ¬P is true and material conditionals are the lower bound, the material conditional $(if\ P)(Q)$ is true, too. Now suppose Q is true. The argument goes just as before, except now we rely on the fact that $Q \wedge P$ entails Q. So if $\neg P \vee Q$ is true so is $(if\ P)(Q)$. □

As before, we either follow the argument to the material conditional or we look for escape routes. That is a tidy way of organizing our choices.

One escape route—flat-footedly denying modus ponens—isn't very popular. Though, as things go in philosophy, that's not because no one has pushed for it. McGee (1985) sees boundedness arguments as pitting modus ponens against import/export. He thinks there are counterexamples to modus ponens and not to import/export, so that forces his choice. (The counterexamples all involve a right-nested indicative we think true (and whose antecedent we think true) even though we do not especially think the embedded conditional on its own is true. There is a not-small literature on the status of the counterexamples.) Others deny some mixture of modus ponens and import/export, but not obviously for reasons connected to boundedness-style arguments (Lycan (2001) denies both import/export and (for good measure) modus ponens, but not without cost).

3 Horseshoe

The *horseshoe theory* is both the simplest response to arguments like the boundedness argument and the simplest story about indicatives: $(if\ \cdot)(\cdot)$ has just the truth conditions that ⊃ does.

Horseshoe Theory Indicatives are material conditionals:
$[\![(if\ P)(Q)]\!]^{c,w} = 1$ iff either P is false or Q is true (at w, in c).

This obviously treats indicatives as truth-functional.

Assuming truth-functionality, is there any other way of assigning truth conditions? No.

Argument. This

(10) If Jimbo is taller than 6 feet, then Jimbo is taller than 5 feet. $(if\ P)(Q)$

should be true (at w) no matter what. Possible heights for Jimbo: (i) over 6 feet (antecedent true), (ii) between 5 and 6 feet (antecedent false, consequent true), and (iii) less than 5 feet (antecedent and consequent false). These are the conditions under which $P \supset Q$ is true. So there's one indicative true when its corresponding horseshoe is. Assuming truth-functionality, this has to be true for *any* indicative. Moreover, indicatives must be false when their corresponding horseshoes are false: otherwise, all indicatives would always be true. So the horseshoe is the only choice. □

The horseshoe theory is, gently put, not the most widely held view these days. But it's not as though it has nothing going for it. Indicatives certainly behave like material conditionals in mathematical contexts. And the horseshoe theory says that instances of the direct argument, import/export, and modus ponens—an impressive who's who in properties conditionals seem to have—strike us as entailments because they are.

It's easy for the horseshoe theory to claim entailments (to conditionals). That is because the material conditional is so weak. The problem is—for just that reason—we have *more* entailments (to conditionals) than we want. (In)famously among them: the paradoxes of material implication.

(11) a. Carl came alone.
??So: if Carl came with Lenny, neither came.
b. Billy got here first.
??So: if Alex got here before Billy, Billy got here first.

These don't seem like entailments even though the truth of either $\neg P$ or Q at w secures the truth of $P \supset Q$ at w. The horseshoe champion owes us some answers.

What we need is an extra-semantic explanation saying that these are entailments, all right, but ones we can live with because the pragmatic facts about conditionals—how they are reasonably and appropriately used—explains their weirdness. That is Grice's (1975) strategy. Conditionals like the one in (11a) are true but ruled out on general grounds. Once we know Carl came alone, it is pointless (though true) to say that either he didn't come with Lenny or neither came. Since it is pointless, as Lewis (1976: p.142) says, "also it is worse than pointless: it is misleading." Ditto for (11b): if you are in a position to say Q then you shouldn't be in the business of saying the weaker conditionalized thing. This kind of explaining away is principled. It doesn't rely on anything special about conditionals: once we take on board the Gricean picture of conversational implicature, we can say something about why the examples in (11) are odd even though the underlying entailments are real. (Lewis (1976) initially goes for this defense of the horseshoe theory for indicatives, using (part of) it to explain why assertability of indicatives goes with their conditional probability and not with the probability of their truth.)

Another kind of explanation: say that indicatives *conventionally* implicate something extra. Jackson (1991) develops this defense, saying that they require "robustness": when you issue an indicative (*if* P)(Q) it conventionally implicates that your credence in? $P \supset Q$ conditional on P—which is to say that the probability of Q conditional on P—is high enough. (Lewis eventually drops the conversational story and instead goes for a variant of Jackson's conventional implicature story.) So *if* is supposed to be like *but:* truth-functional in meaning with an extra, conventionally encoded signal. The problem with, for instance, (11a) is that our reason for thinking the conditional is true is that we think the antecedent false—that's a paradigm case where robustness fails (we drop the conditional upon learning that—to our surprise—Carl came with Lenny). Similarly for (11b). There is little independent evidence of this conventional implicature however (see Bennett 2003: §16 and Edgington 2009: §4.2).

The implicature explanations differ (in mechanism and commitment) but have a lot in common. Both say that it is some extra-semantic fact about using indicatives that explains away the oddness of the entailments in (11). Both say that when it comes to indicatives it's assertability preservation (not some notion of entailment tied to the semantic values of indicatives) that matters and that judgments about entailment may well be clouded by judgments about assertability preservation.

But both also have difficulty explaining facts about indicatives occurring unasserted in embedded environments. The mechanisms for the assert-the-stronger conversational implicature explanation don't apply. And we have no story about how a special-purpose

robustness conventional implicature projects out from under embedding constructions. That is too bad because conditionals *do* occur in larger environments and the horseshoe theory inherits some pretty bad commitments when they do. Everyone knows it was either the gardener, the driver, or the butler. So when the novice on the beat declares that if the gardener's alibi checks out then you can arrest the butler, you rightly object:

(12) a. It's not so that if the gardener didn't do it then the butler did.
b. Just because the gardener didn't do it that doesn't mean the butler did.

In (12a) we have an ordinary indicative under a wide scope sentence negation. (This may be inelegant English, but it's still English.) What you say doesn't entail that the gardener didn't do it. (And it doesn't entail that the butler didn't do it.) This line of reasoning is not good:

(13) Since the driver is still a suspect, it's not so that if the gardener didn't do it then the butler did.
??So: it wasn't the gardener.

The point is that you can be all signed-up for $\neg$ (*if* P)(Q) without being signed-up for the truth of P. But if $\neg(P \supset Q)$ is true then so is P. This is the price material conditionals pay for being so weak: their negations are strong. So we need to explain away this entailment. But the implicature gymnastics invoked thus far don't seem up to it. Maybe some *other* pragmatic mechanism is at work—perhaps the apparent widescope negation isn't a regular negation at all but is really a denial or metalinguistic negation operator (Horn 2001): when you say (13a) you are not asserting the negation of a conditional but you are denying that you will assert the embedded conditional. (This is Grice's reply. Horn seems happy enough to follow him in part because the negation is so awkward—a test, he says, for metalinguistic negation. Merely being resistant to embedding under negation lumps things wrong, though: some modals don't like it (*may* and *must*) even though others don't mind at all (*can* and *have to*). But *may* and *can* (and *must*/*have to*) express the same thing and that thing (whatever it is) is a thing we can sign-up for negating.) This seems dodgy: we're now wheeling in another explaining-away to save our original explaining-away. And the new explaining-away won't help once we embed the (apparent) negated conditional in an even larger environment:

(14) If there is no God, then it's not so that if I pray my prayers will be answered.

So the implicature defense needs work, including saying how the speech act of denial compositionally mixes with the assignment of semantic values for the rest of the language.

4 Variably Strict Conditionals

Assume that indicatives do not have the truth conditions of material conditionals. They say more. What can that more be? Stalnaker (1975) argues that indicatives share a core semantics with counterfactuals. Both kinds of conditionals, he says, are *variably strict* conditionals, but indicatives are governed by additional pragmatic mechanisms that counterfactuals aren't subject to. The variably strict semantics allows indicatives to say more than their horseshoes, and the pragmatic mechanisms explain why certain patterns that *aren't* entailments nevertheless seem so compelling.

Start with orderings over the set of possibilities: for each w, assume there is a (connected, transitive) relation $\leq_w$ recording relative similarity or closeness between worlds. There are two points of interaction between contexts and the ordering: (i) what ordering is relevant is contextually determined and (ii) Stalnaker posits a substantive pragmatic constraint linking contextually relevant possibilities and the ordering. Let's simply model contextually relevant information determined by contexts as functions from worlds to sets of compatible worlds.

(15) a. If there is sugar in the coffee, it tastes sweet.
(if P)(Q)
b. #So: If there is sugar and diesel oil in the coffee, it tastes sweet.
(if (P^ R))(Q)

(16) a. If it rains, it won't pour.
(if P)($\neg Q$)
b. #So: If it pours, it won't rain.
(if Q)($\neg P$)

The nearest P-worlds need not include the nearest (P^R)-worlds, so (15a) won't entail (15b). Similarly, the nearest rainy worlds can all be drizzly worlds,in which case (16a) is true. Still, the nearest pouring-worlds will be rainyworlds and so (16b) will be false.

Since indicatives are stronger than horseshoes, the direct argument isalso classified as a non-entailment:

Counterexample. Suppose w is a ($\neg P$^Q)-world and v is a (P^ $\neg Q$)-world and that v is the closest P-world to w. □

Why are instances of it—like (6)—so compelling? Stalnaker's (1975)answer is that it is a reasonable inference. Put it this way:

Reasonable Inference Suppose P is successfully asserted (at w in c) and c' is the resulting posterior context. P, so : Q is a reasonable inference iff Q is accepted in c'.

This is not a semantic property. It depends on what speakers are up to in asserting various things. In order to be successfully asserted, a premise has to first be felicitously asserted. In order for a conclusion to be accepted in a context, it has to get a clean bill of pragmatic health. Since or-to-if begins with a disjunctive premise and ends with a conditional conclusion, we need two additional pieces of information. Begin with the disjunctive premise: in order for a disjunction to be felicitously asserted in a context either disjunct might be true without the other. That seems plausible.

The constraint on indicatives is more substantive. In order to use an indicative conditional (owing perhaps to their epistemic connection) there is an additional requirement that the selected antecedent world(s) must be compatible with the context, assuming the antecedent is:

Variably Strict Semantics Indicatives are variably strict conditionals:
Assume for any worlds w, v and proposition X, that $\leq_w$ is such that:

i. (Centering) w is minimal in $\leq_w$: if $v \leq_w$, w then $w = v$
ii. (Limit) X has at least one $\leq_w$-minimal world
iii. (Uniqueness) X has no more than one $\leq_w$-minimal world
Then $[\![(\textit{if } P)(Q)]\!]^{c,w} = 1$ iff Q is true at the $\leq_w$-minimal P-world.

This is Stalnaker's (1968) set-up. (Lewis's (1973) version differs by making neither the Limit Assumption nor the Uniqueness Assumption. The pragmatic mechanisms for indicatives that Stalnaker develops are independent of whether the core semantics is Stalnakerian or Lewisian. I'll make the Limit Assumption and be willfully sloppy about whether what's required is the closest antecedent world or the set of closest antecedent worlds.)

This predicts that indicatives say more than their corresponding horseshoes. The paradoxes of material implication won't be reproduced: the inferences in (11) are non-entailments for a variably strict conditional. And while the material conditional validates antecedent strengthening and contraposition, variably strict conditionals don't:

Pragmatic Constraint If P is compatible with the context c, then the $\leq_w$-minimal P-world(s) is compatible with c.

This is motivated this way. First: indicatives are happiest being asserted when their antecedents might, in view of the context, be. If you want to say something conditional on P in a context in which it is settled that $\neg P$, then the counterfactual is what you need to reach for. Second: in asserting indicatives, you are trying to say something about the possibilities compatible with the context. So the worlds relevant to whether an indicative is true had better be compatible with the context. This plus the plausible requirement for disjunctions guarantees that the direct argument is a reasonable inference:

Argument. Suppose an assertion of (6a) is successful (felicitous and accepted) at w in a context c. So $c(w)$ contains some $(P \wedge \neg Q)$-worlds and some $(\neg P \wedge Q)$-worlds. Since the disjunction is then accepted, we eliminate worlds from $c(w)$ where it isn't true. In the posterior $c'(w)$, there are only $(P \vee Q)$-worlds. Some are $(P \wedge \neg Q)$-worlds and some are $(\neg P \wedge Q)$-worlds. Now consider the indicative (6b): (*if* $\neg P$)(Q). Its antecedent is compatible with c' (there are $\neg P$-worlds in $c'(w)$). Thus the pragmatic constraint ensures that the ($\leq_w$-)closest $\neg P$-world(s) are also in $c'(w)$. But there aren't any $(\neg P \wedge \neg Q)$-worlds in $c'(w)$. So the closest $\neg P$-worlds will be Q-worlds. So the indicative is accepted in c'. □

That's why, according to Stalnaker, *or*-to-*if* instances—though invalid—are so compelling.

A variably strict semantics also fails to classify pairs of conditionals like (7) as mutual entailments: whatever the merits of import/export, given the variably strict semantics, it ain't a validity. That's because the nearest Q-world to the nearest P-world to w need not be the same as the nearest $(P \wedge Q)$-world to w, even when both P and Q are compatible with the context. In the case of the direct argument, there is an independently plausible felicity-requirement on the disjunctive premise—that in the prior context either disjunct may be true without the other—that does real work in making sure the posterior context will be one in which the indicative is accepted. For import/export it's not obvious what such an extra, independently plausible felicity-requirement on the conditional premises (in either direction) would look like. And so it's not obvious how to use the concept of reasonable inference to explain why import/export instances—though invalid—are so compelling.

5 NTV and Conditional Assertion

The boundedness argument can be seen as an argument that there is no good way of assigning truth conditions to indicatives: the only options come up short either by giving up something important or by saddling us with the horseshoe theory. Gibbard (1981) sees it that way and sees it as reason conclude, with Adams (1975), that indicative conditionals don't have truth conditions at all: they do not report conditional information but are a means of expressing conditional belief. Edgington (1995, 2009) and Bennett (2003) (among others) follow them to this "N(o)T(ruth)V(alue)" conclusion. (This view sounds like it denies that indicatives have truth values. In that it does not disappoint. But the spirit is to deny something more. We've assumed that conditionals have the same sort of semantic value that other bits of declarative language have. NTVers deny that indicative conditionals have or traffic in the same kind of semantic unit of exchange that other declarative parts of the language do: it's not truth and it's not some richer-than-truth thing either.)

Another argument against truth conditions, this one due to Edgington (2009: §2): truth conditions come in two flavors (truth-functional and non-truth-functional) and neither seems right for indicatives. Truth-functional truth conditions are out because the horseshoe is out. That leaves non-truth-functional ones. Saying that the truth conditions for (*if P*)(*Q*) are non-truth-functional means that the truth values of *P* and *Q* at *w* don't fix the truth value of (*if P*)(*Q*) at *w*. So if *P* is false at *w* and *Q* is true at *w* then sometimes (*if P*)(*Q*) is true at *w* and sometimes it isn't. That's *variability*. But we have the *or*-to-*if* inference: if all the information you have is that $\neg P \vee Q$ then that is *always* sufficient for (*if P*)(*Q*). This is *uniformity*. Edgington says variability and uniformity are on a collision course and no non-truth-functional theory can be squared with this. If all that is right, we are out of options truth conditions-wise. (I myself think this isn't all right, since for all that's been said it's possible that the variability be variability in truth value at a given world between contexts and the uniformity be uniformity across worlds compatible with a given context (Gillies 2009: §4).)

A third argument is based on hedging. Lewis (1976, 1986) showed that on pain of triviality there are no conditional propositions such that their probability of truth always equals the conditional probabilities of their consequents given their antecedents. (The basic result has been fortified and extended: see, e.g., Hájek (1994). Bennett (2003: §25–31) surveys some of the main territory.) There is a connection that needs explaining:

(17) a. Probably, if my team doesn't sign some big players they won't win.
b. If my team doesn't sign some big players they won't win.

The hedged (17a) seems to say that the probability that my team will continue their losing ways, conditional on them continuing their frugal/mismanaging ways, is high enough. I don't think that they definitely won't win if they don't make some big moves, but I think that's likely (enough). So the degree to which I am signed up for the bare (17b) seems to track what (17a) says. Suppose *B* is a probability function representing degrees of belief or assertability or whatever. That's some reason to sign-up for this:

The Equation $B((\textit{if } P)(Q)) = B\text{-plus-}P(Q)$

We haven't yet said anything about how we arrive at this posterior probability *B*-plus-*P*(*Q*). Maybe it's through conditionalization, maybe not. The Equation implies:

Linearity If *B* is a mixture of B_1 and B_2 then *B*-plus-*P* is a mixture of B_1-plus-*P* and B_2-plus-*P*.

Proof. Assume *B* is a mixture of B_1 and B_2. Consider any *P* and Q. So $B((\textit{if } P)(Q))$ is a mixture of $B_1((\textit{if } P)(Q))$ and $B_2((\textit{if } P)(Q))$. But since we're assuming The Equation that means *B*-plus-*P*(*Q*) is a mixture of B_1-plus-*P*(*Q*) and B_2-plus-*P*(*Q*). □

Gärdenfors (1982) showed that Linearity is at odds with conditionalization. So The Equation implies that the *plus* in *B*-plus-*P* can't be conditionalization:

Observation The Equation is not compatible with conditionalization.

One way of seeing things: Gärdenfors showed that Linearity exactly characterizes changing the *B*'s by *imaging* (imagining and conditionalization can agree only at the boundaries). Another way: conditionalization implies *conservativism:* whenever $B(Q)=1$ and $B(P) > 0$, then $B_P(Q) = 1$. Linearity rules that out (modulo triviality). Lewis's (1976) main triviality results then follow as corollaries.

(All of this is replicated if we think of the *B*'s as sets of full beliefs. The qualitative counterpart of The Equation:

(RT) $(\textit{if } P)(Q) \in B$ iff $Q \in B * P$

Again, no restrictions on how we arrive at the posterior $B{*}P$. (RT) implies the qualitative counterpart to Linearity:

(Mon) $B_1 \subseteq B_2$ implies $B_1 * P \subseteq B_2 * P$.

(Suppose $Q \in B_1 * P$. Then $(\textit{if } P)(Q) \in B_1$ and so $(\textit{if } P)(Q) \in B_2$. So $Q \in B_2 * P$.) Just as Linearity is at odds with conservativity, (Mon) is at odds with it's counterpart "preservation" ($\neg P \notin B$ implies $B \subseteq B * P$) (Gärdenfors 1988; Segerberg 1989).)

There are nearby problems, too. I'm dead certain that the Cubs didn't win it all this year, but think it's possible that they won at least 70 games (at a certain point, you have to just quit keeping track).

(18) a. ??Probably, if the Cubs won at least 70 games, then they won it all this year.
b. If the Cubs won at least 70 games, then they won it all this year.

Given what I know, there's no hedge low enough for (18a) to be OK and no degree of belief small enough for me to sign up for the bare (18b) even to that minimal degree. Bradley (2000) argues that treating indicatives as things expressing propositions—subject to the same probability hedges that non-conditional stuff is—means having to say that it is sometimes OK to give positive weight to a conditional and its antecedent even though you give no weight at all to its consequent.

So if indicatives express propositions of the normal sort and scope under probability operators, we have trouble. NTVers say so much the worse for the idea that indicatives

express propositions, have truth conditions, and determine truth-values at all. The reason you sign-up for an indicative just to the degree that you conditionally sign-up for its consequent given its antecedent is the straightforward reason that indicatives are vehicles expressing (not representing) such conditional signings-up.

NTVers don't have to worry about triviality under hedges, but this is an open route only if they can explain why inferences involving indicatives seem so compelling—it can't be because they are entailments, the preservation of the main semantic value from premises to conclusion or something like that. Here they have something to say: our intuitive judgments are tracking Adams's (1975) notion of *probabilistic validity*—an argument is probabilistically valid iff the uncertainty of its conclusion is no more than the sum of the uncertainties of its premises (for an introduction see Adams (1998)). Modus ponens and modus tollens are probabilistically valid, the paradoxes of material implication (11) aren't. But neither is conditional proof. And without saying something extra and special-purpose about embedded conditionals, neither is import/export.

Embedding is another worry for NTVers. That is because they have to worry about what we're doing when we issue ordinary indicatives. Conditionals aren't, they say, for asserting. They are vehicles of conditional assertion. So in uttering an indicative (*if P*)(*Q*) (at *w*, in *c*), we are conditionally asserting *Q*-on-*P* in *c* at *w*. That affects *c*(*w*) in the same way that flat-out asserting *Q* affects *c*(*w*)-plus-*P*. But for this to fly, *Q* has to be the sort of thing that can be flat-out asserted in the first place. That means NTVers have revisionary work to do to explain away apparently negated conditionals like (12) and revisionary work to do to explain away the apparently right-nested indicatives featured by import/export. They also face hard choices when it comes to epistemic modals. Consider:

(19) a. If Red isn't in the box, Blue must be.
b. If Red isn't in the box, Blue might be.

Either we extend the NTV thesis to the modals or not. If so, then embeddings like (19) are prima facie ruled out. If not, we cut ties between the modals and conditionals. Sometimes NTVers say that embedding facts go the other way: since indicatives don't always freely embed (they are tough to negate and it's tough to left-nest them) this suggests that they don't express propositions. This is too fast. Some lexical items have syntactic restrictions on embedding that have nothing to do with their semantics. We've seen this already: *might* and *must* resist embedding under negation (and resist embedding under deontic modal operators) even though *can* and *have to* aren't so fickle.

6 Epistemic Operators

Set aside NTV views. Interaction between indicative conditionals and epistemic operators (hedges and modals alike) is still tricky. Examples:

(20) a. If the gardener isn't guilty, the butler must be.
b. If he didn't tell Harry, he probably told Tom.
c. If Carl is here, then presumably Lenny is here.
d. If the Cubs get good pitching and timely hitting, they might win.

These operators seem to occur embedded, giving us (apparently) instances of this:

(21) (*if P*)(OPERATOR *Q*)

This takes *if*-of-English to contribute a conditional relation as its meaning, and further says that that conditional relation holds between *P*-worlds and worlds where OPERATOR *Q* is true. The trouble is there seems to be no such conditional relation. Grant some minimal assumptions about what conditional relation between (relevant) antecedent-worlds and consequent worlds an (*if* ·)(·) picks out (at w in c). (For instance, the relation has to be idempotent and right upward monotonic and it has to care about consequents.) Then:

Observation The only conditional relation that (*if P*)(*Q*) can express between the relevant antecedent worlds and consequent worlds is $\subseteq$.

For proofs (of slightly different versions of this), see van Benthem 1986; Veltman 1985; Gillies 2010.

Now grant some minimal assumptions about the *if*-relevant worlds at w: (i) w is always relevant to an indicative (*if P*)(*Q*) at w; (ii) the *if*-relevant worlds (in c at w) are compatible with c. And grant that modals are quantifiers over worlds compatible with the context: a *might* at w in c is an existential quantifier over $c(w)$ and *must* is the dual universal quantifier.

To highlight the trouble: I have lost my marbles and know that one and only one—either Red or Yellow—is in the box.

(22) a. Red might be in the box and Yellow might be in the box.
might P $\wedge$ *might Q*

b. If Yellow isn't in the box, then Red must be.
(*if* $\neg$ *Q*)(*must P*)

c. If Red isn't in the box, then Yellow must be.
(*if* $\neg$ *P*)(*must Q*)

These are all true. But it seems they can't all be true together if (*if* ·)(·) means $\subseteq$. No matter what semantics for indicatives we pick, the sentences in (22) are incompatible. While it's a lamentable fact that my marbles are lost, it's still a fact not an impossibility.

Proof Sketch. Suppose otherwise and that we have just two worlds compatible with the context, w and v. Look at w. it's either a $\neg Q$-world or a $\neg P$-world. Suppose it's a $\neg Q$-world. Since (*if* $\neg Q$)(*must P*) is true at w in c all the *if*-relevant $\neg Q$-worlds are worlds where *must P* is true. But w must be one of the relevant *if*-relevant worlds and it's a $\neg Q$-world, too. So it's a world where *must P* is true. So $c(w)$ has only *P*-worlds compatible with it. But *might Q* is true at w! Mutatis mutandis if w is a $\neg P$-world. (This is equally an apparent argument against modus ponens for indicatives, assuming that entailment is simple preservation of truth(-at-a-point). □

Ah, the problem is that we got the scope relations wrong! Epistemic operators, you say, scope over the indicatives in environments like (20). Replace (21) with

(23) OPERATOR (*if P*)(*Q*)

This is no better, really. For one thing if OPERATOR is a probability hedge, triviality looms. (There are other reasons, but they involve questions about whether we decide to be *egalitarian* (all worlds compatible with the context are created equal when it comes to saying which are *if*-relevant) or *chauvinistic* (not egalitarian). Those questions are left open given what we've said, but when paired with the widescoping strategy in (23), each leads to different trouble (Gillies 2010: §6).) We appear to be out of choices, scope-wise.

7 Restrictor View

This trouble is a lot like some trouble Lewis (1975) saw for *if*s in quantificational environments like these:

(24) { Always / Sometimes / Seldom } if a farmer owns a donkey, he beats it.

What single connective could *if* contribute in each of these? Maybe something iffy would work if the adverb is the universal *always*, but conjunction would be better if its an existential like *sometimes* and neither looks good for *seldom*. The *if* has the job of restricting the quantifiers. Lewis's (1975) conclusion: the restricting job isn't one that a conditional operator can do. So sentences like those in (24) are not instances of a conditional operator plus an adverb of quantification. Instead, he said, the *if*s are a *nonconnective* whose only job is to mark an argument slot for the adverb of quantification.

Kratzer's idea (developed in, e.g., Kratzer 1981, 1986, 1991, 2010) is that this holds not just for *if*s under adverbs of quantification but for *if*s across the board. The thing all *if*s do is restrict operators. So they aren't iffy. The choice between (21) and (23) is not what we are after. Instead what we have is something like this:

Restrictor Analysis Conditionals are restricted operators with logical forms like this:
Quantifier/Operator + *if-clause* + consequent clause
OPERATOR(P)(Q)

When the operator is a modal like *must* or *probably* or *might*, the job of the *if*-clause is to restrict the domain over which the modal quantifies. If there is no overt operator, then since *if*s restrict operators, there must be a covert (necessity) one.

This picture is general and powerful. (For more on the restrictor view, see von Fintel (2011) and the references therein.) It is, for instance, easy to see why indicatives go in for import/export. It is also easy to see how and why conditionals (which is to say "conditionals") and operators—like the unrestricted *might*s in (22)—interact. This is true for fancier modals and hedges, too:

(25) a. If he's a Quaker, he's presumably a pacifist.
b. If the bet is on odd, it's probably a loser.
c. If the coin is fair, then the probability of heads is ½.

Each of these expresses some restricted hedge: that he's presumably a pacifist (given he's a quaker), that the bet won't be won (given it's on odd), that the probability of heads is

$\frac{1}{2}$ (given the coin is fair). The hedges—*presumably*, *probably*, and *x-probably*—are modal operators. Treating *if*s as restrictors means that when we look at environments like these we shift our attention to the operators involved. Getting straight about them is how we get straight about the conditionals in which they occur. Though our current topic is indicatives and not hedges like these, I'll just briefly mention two routes to modeling their contribution that fit hand/glove-wise with the restrictor view of *if*s.

Route one: introduce a qualitative ordering between possibilities and eek truth-conditions for a hedge at w in c out of this. This is the route that Kratzer takes. (There are a lot of choices for the qualitative structure used and choices for how to connect it to the hedges. Yalcin (2010) surveys some of the choices.) Route two: introduce a more fine-grained body of information at hand determined by a context c. In particular: allow contexts to provide probability measures over the possibilities compatible with them. (Here, too, there are lots of choices.) Each route ties hedges to a certain body of information and that body of information can be restricted in pretty straightforward ways: all the choices allow for a well-understood way of "updating" information that bears a systematic and well-understood connection to conditionalization. (For more on the choices for representing the needed uncertainty the hedges trade on, and how those ways of modeling it go in for updating, see Halpern 2003.)

8 Re-Inventing Conditional Connectives

There are still reasons to explore properly iffy operators that carry conditional meanings. One reason: the charge is that the restricting job that *if*s unquestionably do is a job that can't be done by any conditional operator. Does the charge stick? It's worth knowing. (Another reason: the way hedged *if*s (and their partnering iffed hedges) behave in conversation seems to require a conditional connective to figure in the story. (The history of the problem is a little involved, but see von Fintel & Gillies (2011).) That would be surprising not least because Lewis-style triviality results seem to push exactly the other way. But the jury is currently still out on this particular bundle of puzzles.) The re-inventions sketched here also have an answer on offer for what the extra bit is that indicatives say over and above their corresponding horseshoes (naturally enough the bit the different re-inventions offer do differ).

One way that *if* can express a uniform conditional meaning while still doing its restricting job is for the conditional meaning to be a gappy one—gaps for when things aren't as the *if*-clause says—and for our story of embedding operators to be sensitive to those gaps. The idea is Belnap's (1970) and it has been revived and pushed by Huitink (2009). (Lewis knew this was a way out, but dismissed it.)

Here is the bare bones Belnap view: a conditional (*if* P)(Q) (in c at w) says pretty much the same as Q does (in c at w) provided P is true. Otherwise, it says nothing. Two ways to try to make this go: "boring" and "interesting". The boring way says sameness is sameness of *truth-value*, interesting that it is sameness of *semantic value*. Interestingly, it is the boring way that seems more promising. This is what Huitink does.

Belnap Conditional Indicatives express gappy propositions:

$$[\![(\textit{if } P)(Q)]\!]^{c,w} = \begin{cases} [\![Q]\!]^{c,w} & \text{if } [\![P]\!]^{c,w} = 1 \\ \text{undefined} & \text{otherwise} \end{cases}$$

This treats (*if P*)(*Q*) like the horseshoe when *P* is true but diverges from it in just the cases that the horseshoe theory goes wild. But it doesn't by itself meet the challenge of getting the restricting behavior right. If the operator that needs restricting is out-scoped by the *if* we will get the wrong and unrestricted truth-conditions and if the operator does the out-scoping we will *still* get the wrong and truth-conditions. The fix is easy. First: widescope the operators. Second: reign in their quantificational domains to survey only worlds/cases/situations/whatever at which the embedded sentence gets a truth-value. That's it.

There is another way to re-invent conditional connectives based on Ramsey's (1929/1990) suggestion. The Ramsey test is a recipe for when to *accept* a conditional: you accept (*if P*)(*Q*) in a state *B* iff *Q* is accepted in the subordinate state got by taking *B* and adding the information that *P* to it. But we want a story about what *if*s mean, not so much about when they are/ought to be accepted. Again the fix is easy: (*if P*)(*Q*) in a context *c* says that all (relevant) *P*-possibilities are possibilities at which *Q* is true. What is the context relevant for checking at those *P*-possibilities whether *Q* is true? The subordinate context got by taking *c* and adding the information that *P* to it. Making the straightforward choice for how the adding goes (zoom in to the possibilities where *P* is true):

Shifty Conditional Indicatives express shifty propositions:
$[\![(\textit{if } P)(Q)]\!]^{c,w} = 1$ iff $c(w) \cap [\![P]\!]^c \subseteq [\![Q]\!]^{c+P}$
where $c + P = \lambda v.c(v) \cap [\![P]\!]^c$

If we narrowscope the relevant operators, leaving their plain vanilla semantics intact, this is enough to get the restricting behavior of *if*-clauses to mesh with them expressing a genuine conditional connective.

This re-invention is equivalent to dynamic semantic accounts that types all sentences in the language as programs or instructions for changing the context. (The classic references: Groenendijk Stokhof (1991); Veltman (1996).) Intuition: a program means what it does, and what it does depends on what things are like—the state you are in—when you execute it. So, a program's content is a relation between prior states and corresponding posterior states.

Dynamic semantics treats all sentences that way: their contents are the characteristic changes that successful assertions of them induce. A simple example:

Dynamic Semantics for Propositional Logic Take a state (context) *s* to be a set of worlds.
i. $s[P] = \{w \in s$: *P* is true in *w*$\}$ for atomic *P*'s
ii. $s[\neg P] = s \setminus s[P]$
iii. $s[P \wedge Q] = s[P][Q]$

Here the relation is in fact a function [·] from states to states (read it postfix). So (successfully asserting) atomic sentences tell us to throw away worlds where they aren't true, negations tell us to throw out what would survive an update with the things negated, and conjunctions tell us to process things in order. The point of each these instructions is to have a non-null upshot.

Not all instructions have that aim. You say to me: *Check whether the game is on*. You're not instructing me to change anything game-wise or TV-wise but instructing me to see whether the state we are in has a certain property. It's a *test* program. The dynamic idea is that indicatives are Ramsey-inspired test instructions:

Dynamic Conditional Indicatives express test programs:
$s[(if\ P)(Q)] = \{w \in s: s[P][Q] = s[P]\}$

Let's say that a sentence P is true in state s iff the information that P is already present in s. That is: P is true in s iff $s[P] = s$. Then we can put things this way: an indicative (*if P*) (*Q*) is true in a state s iff Q is true in the subordinate state $s[P]$ got by taking s and adding the information P to it. That is very Ramseylike.

Acknowledgments

I've learned a lot about conditionals from a lot of people. Some of my personal intellectual debts are reflected in the references, some not (or not enough). So my biggest debtholders: Nicholas Asher, Josh Dever, Frank Veltman, and, of course, my frequent collaborator Kai von Fintel. Finally, I owe a tremendous debt to John L. Pollock. I'd like to dedicate this to his memory. He will be missed (unconditionally).

References

Adams, Ernest W. 1975. *The Logic of Conditionals*. D. Reidel.
Adams, Ernest W. 1998. *A Primer of Probability Logic*. Center for the Study of Language and Information.
Anderson, Alan Ross. 1951. A note on subjunctive and counterfactual conditionals. *Analysis* 12(2). 35.
Austin, John L. 1956. *Ifs* and *cans*. In *Proceedings of the British Academy*, 109–132.
Belnap, Nuel D. 1970. Conditional assertion and restricted quantification. *Noûs* 4(1). 1–12.
Bennett, Jonathon. 2003. *A Philosophical Guide to Conditionals*. Oxford University Press.
van Benthem, Johan. 1986. *Essays in Logical Semantics*. D. Reidel.
Bradley, Richard. 2000. A preservation condition for conditionals. *Analysis* 60(3). 219–222.
Edgington, Dorothy. 1995. On conditionals. *Mind* 104(414). 235–329.
Edgington, Dorothy. 2009. Conditionals. In Ed Zalta (ed.), *Stanford Encyclopedia of Philosophy*, vol. Spring 2009 Edition. URL http://plato.Stanford.cdu/archives/ win2003/entries/conditionals/.
von Fintel, Kai. 1998. The presupposition of subjunctive conditionals. In U. Sauerland & O. Percus (eds.), *MIT Working Papers in Linguistics: The Interpretive Tract*, vol. 25, 29–44.
von Fintel, Kai & Anthony S. Gillies. 2011. Hedging your *If*s and vice versa. Manuscript, MIT/Rutgers.
van Fraassen, Bas C. 1980. Review of Brian Ellis, *Rational Belief Systems*. *Canadian Journal of Philosophy* 10. 497–511.
Gärdenfors, Peter. 1982. Imaging and conditionalization. *The Journal of Philosophy* 79. 747–760.
Gärdenfors, Peter. 1988. *Knowledge in Flux*. MIT Press.
Gibbard, Allan. 1981. Two recent theories of conditionals. In W. L. Harper, R. Stalnaker & G. Pearce (eds.), *Ifs*, 211–248. D. Reidel.
Gillies, Anthony S. 2009. On truth-conditions for *If* (but not quite only *If*). *The Philosophical Review* 118(3). 325–348. doi:10.1215/00318108–2009–002.
Gillies, Anthony S. 2010, Iffiness. *Semantics and Pragmatics* 3(4). 1–42. doi:10.3765/sp.3.4.
Grice, H. Paul. 1975. Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Syntax and Semantics*. Academic Press.
Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14. 39–100.
Hájek, Alan. 1994. Triviality on the cheap? In Ellery Eells & Brian Skyrms (eds.), *Probability and Conditionals: Belief Revision and Rational Decision*, 113–140. Cambridge University Press.
Halpern, Joseph Y. 2003. *Reasoning about Uncertainty*. MIT Press.
Heim, Irene. 1983. On the projection problem for presuppositions. *WCCFL* 2. 114–125.
Huitink, Janneke. 2009. *Domain Restriction by Conditional Connectives*. Goethe-University Frankfurt.
—— (2010c) "What 'Intuitions' are Linguistic Evidence?" Erkenntnis 73: 251–64.
—— (2010d) "Linguistic Intuitions Revisited," British Journal for the Philosophy of Science 61: 833–65.
—— (2011a) "Experimental Semantics," Philosophy and Phenomenological Research LXXXII: 418–35.
—— (2011b) "Methodology and the Nature of Knowing How," Journal of Philosophy CVIII: 205–18.

Horn, Laurence R. 2001. *A Natural History of Negation*. CSLI Publications, reissue edition edn.
Jackson, Frank. 1991. *Conditionals*. Oxford University Press, USA.
Kratzer, Angelika. 1981. The notional category of modality. In H. J. Eikmeyer & H. Rieser (eds.), *Words, worlds, and contexts, new approaches in word semantics*, 38–74. Berlin: de Gruyter.
Kratzer, Angelika. 1986. Conditionals. *Chicago Linguistics Society* 22(2). 1–15.
Kratzer, Angelika. 1991. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 639–650. Berlin: de Gruyter.
Kratzer, Angelika. 2010. *Papers on Modals and Conditionals*. Oxford University Press.
Lewis, David. 1973. *Counterfactuals*. Blackwell.
Lewis, David. 1975. Adverbs of quantification. In E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge University Press.
Lewis, David. 1976. Probabilities of conditionals and conditional probabilities. *The Philosophical Review* 85. 297–315.
Lewis, David. 1986. Probabilities of conditionals and conditional probabilities II. *The Philosophical Review* 95. 581–589.
Lycan, W. G. 2001. *Real Conditionals*. Oxford University Press.
McGee, Vann. 1985. A counterexample to modus ponens. *Journal of Philosophy* 82(9). 462–471.
Ramsey, Frank P. 1929/1990. General propositions and causality. In H. A. Mellor (ed.), *F. Ramsey, Philosophical Papers*. Cambridge University Press.
Segerberg, Krister. 1989. A note on an impossibility theorem of Gärdenfors. *Noûs* 23. 51–354.
Stalnaker, Robert. 1968. A theory of conditionals. In Nicholas Reseller (ed.), *Studies in Logical Theory*, 98–112. Basil Blackwell.
Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5. 269–286.
Veltman, Frank. 1985. Logics for Conditionals. Ph.D. thesis, University of Amsterdam.
Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25. 221–261. doi:10.1007/BF00248150.
Yalcin, Seth. 2010. Probability operators. URL http://semanticsarchive.net/Archive/Tg3OTQyZ/yalcin.po.2010.draft.pdf.

3.13 SUBJUNCTIVE CONDITIONALS

Kai von Fintel

1 Introduction

Conditional sentences, canonically of the form "if p, q", whisk us away to a scenario described by their antecedent and then make a claim about it in their consequent. There are two main kinds of conditionals, as illustrated in the well-known Oswald/Kennedy minimal pair (due to Adams 1970):

(1) If Oswald didn't kill Kennedy, someone else did.
(2) If Oswald hadn't killed Kennedy, someone else would have.

Clearly, the two conditionals differ in meaning. The conditional in (1) signals that it is an open possibility that Oswald didn't kill Kennedy and will be judged true by anyone who knows that Kennedy was in fact assassinated. The conditional in (2), in contrast, signals that it is taken for granted that Oswald did in fact kill Kennedy and makes the somewhat dubious claim that Kennedy's assassination was inevitable, perhaps based on a vast conspiracy. The same difference in meaning can be illustrated with a similar pair (due to Bennett or Stalnaker?), where it is even harder to hear the second conditional as making a plausible claim:

(3) If Shakespeare didn't write *Hamlet*, someone else did.
(4) If Shakespeare hadn't written *Hamlet*, someone else would have.

This chapter will be concerned with the meaning of conditionals of the second kind.

Conditionals of the first kind are usually called "indicative" conditionals, while conditionals of the second kind are called "subjunctive" or "counterfactual" conditionals. The "indicative" vs. "subjunctive" terminology suggests that the distinction is based in grammatical mood, while the term "counterfactual" suggests that the second kind deals with a contrary-to-fact assumption. Neither terminology is entirely accurate.

It is clear that the outward difference between conditionals of the two kinds lies in their tense/mood/aspect syntax, but it is not reliably a difference in indicative vs. subjunctive mood properly speaking. In languages that have a past subjunctive (such as German), antecedents of conditionals of the second kind do indeed appear in the subjunctive mood but if a language does not have a past subjunctive, some other form is

used. English uses an indicative antecedent with an additional layer of past tense morphology ("if Oswald didn't kill Kennedy" becomes "if Oswald hadn't killed Kennedy"). It is that additional layer of past tense morphology (which doesn't obviously contribute an actual past meaning) that quite reliably signals conditionals of the second kind. So, instead of calling them subjunctive conditionals, one might call them "additional past" conditionals.

While our initial examples did appear to signal that their antecedent was false, conditionals of the second kind do not always carry the suggestion of counterfactuality:

(5) If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show. (Anderson 1951)

A doctor who utters (5) might be prepared to use it as part of an argument that the antecedent is in fact true, so the conditional could not be conveying counter-factuality as part of its meaning.

Conversely, there are some formally indicative conditionals that express counterfactuality:

(6) If he has solved this problem, I'm the Queen of England.
(7) If Messi waits just a second longer, he scores on that play.

[Indicative counterfactuals of the kind in (7) are common in sportscast play-by-play commentary. They have not (yet) been studied in the semantic literature.]

Even though inaccurate, the terms indicative/subjunctive/counterfactual are so entrenched that it is presumably futile to try to reform usage. The association of conditionals of the second kind with counterfactuality is somewhat tenuous. In comparison, the connection of the distinction to tense/mood/aspect morphology is undeniable even if the morphology in conditionals of the second kind isn't strictly speaking subjunctive in many languages. So, the indicative vs. subjunctive terminology is the slightly less inappropriate one and will be the one we adopt here.

2 The Meaning of Subjunctives

Conditionals whisk us away to a scenario where the antecedent holds. The most influential semantic theories about subjunctive conditionals, in particular the pioneering works of Stalnaker and Lewis, treat them as making claims about "possible worlds" that may be quite different from the actual world. Nevertheless, when one looks at examples like the Oswald/Kennedy conditionals, one sees that even the counterfactual version makes palpable claims about reality: "if Oswald hadn't killed Kennedy, someone else would have" makes the claim that in actuality, Kennedy's assassination was inevitable, perhaps because there was a vast conspiracy in place or because one thinks that Kennedy's actions and policies inexorably provoked an assassination. How does talk about other possible worlds connect to the actual world?

All of the possible worlds semantic approaches answer this concern by closely tying the identification of the relevant set of possible worlds to facts in the actual world. Stalnaker 1968 grounds the basic idea in the same "Ramsey Test" that has inspired accounts of indicative conditionals. For Stalnaker, the test describes how we evaluate conditionals: we hypothetically add the antecedent to our information state and assess

the consequent in the resulting state. While canonical indicatives have antecedents that are compatible with the relevant information state, canonical subjunctives involve antecedents that contradict some prior information. Thus, the adding of the antecedent will necessitate some revisions of the information state. (It should be noted that the proper analysis of Anderson-type examples, as in (5), in this picture is not entirely obvious. See von Fintel 1998 for some discussion.)

Correlated with this pyschological process of assessing a conditional, Stalnaker proposes a truth-conditional semantics that starts from the actual world, considers the antecedent, and looks for worlds that differ minimally from the actual world while making the antecedent true. It is in those worlds that the consequent is then evaluated. So, in the Oswald/Kennedy case, we keep fixed all actual facts that are not strictly tied to Oswald killing Kennedy. The subjunctive conditional claiming that if Oswald hadn't killed Kennedy, someone else would have, thus amounts to saying that there were facts in the actual world that would have led to Kennedy's assassination one way or the other.

Tying the semantics of conditionals to the notion of comparative similarity between possible worlds may seem like it doesn't help us nail down the meaning of conditionals all that much, since similarity obviously is quite a vague and context-dependent notion. In a brilliant and influential move, Lewis (1973: 91ff) defended the use of this notion by pointing out that conditionals themselves are inherently vague and context dependent. Therefore a proper analysis of conditionals needs to correctly reflect their vagueness and context dependence. He argued that the comparative similarity relation between possible worlds is just the right tool to do so.

As an illustration of the vagaries of judging similarity and difference between worlds, consider Quine's famous pair of conditionals (1960: 221; NB: he actually calls the pair "Goodman's, nearly enough"):

(8) If Caesar were in command, he would use the atom bomb.
(9) If Caesar were in command, he would use catapults.

Quine wrote: "we feign belief in the antecedent and see how convincing we then find the consequent. What traits of the real world to suppose preserved in the feigned world of the contrary-to-fact antecedent can be guessed only from a sympathetic sense of the fabulist's likely purpose in spinning his fable." His "preserving of traits of the real world in the feigned world of the antecedent" corresponds quite directly with the decision about which properties of worlds carry more weight in the judgment of similarity between worlds. In (8), we seem to keep constant Caesar's ruthlessness, while in (9) we (also) care about the technologies that were actually at his disposal in his time.

The similarity/difference-based semantics remains quite a schematic framework, to be filled in with contextually relevant considerations for assessing similarities and differences. In later work, in response to several worries about the account, Lewis and others suggested constraints on the kind of similarity relation underlying conditionals. This debate continues rather unabated.

A variant of the semantics based on a similarity ordering is given by "premise semantics", inspired by Goodman (1947) and Rescher (1964) and developed in rival forms by Kratzer (1977, 1979, 1981) and Veltman (1976). Here, we start with a description of the actual world (a set of sentences or propositions) and revise it to make it accommodate the addition of the antecedent. Then, we check whether the consequent follows from the resulting set of premises. The revision process in this account intuitively parallels

the determination of the antecedent worlds most similar to the actual world in the Stalnaker/Lewis framesvork. And in fact, Lewis (1981) showed that technically the two approaches are intertranslateable. More recent work in the premise semantic tradition includes Kratzer 1989, Veltman 2005, Kanazawa et al. 2005, Kratzer 2005.

3 Nonmonotonicity

There is an austere way of spelling out a possible worlds analysis of conditionals according to which the context supplies us with a set of relevant possible worlds, those that in all relevant respects are similar enough to the actual world, and that the conditional then makes a claim about all of the antecedent worlds in that set. This is known as "strict implication". In contrast, the analysis proposed by Stalnaker and Lewis is a "variably strict" analysis (terminology due to Lewis), in which we start from the antecedent and identify among the antecedent worlds those worlds that are as similar as possible to the actual world. The variably strict semantics in distinction to the strict implication analysis predicts a variety of nonmonotonic behaviors, and Stalnaker and Lewis argued that those predictions are borne out.

Under a Strict implication analysis, the pattern known as Strengthening the Antecedent, is predicted to be valid:

(10) *Strengthening the Antecedent*
if p, q $\Rightarrow$ if p&r, q

If all (contextually relevant) p-worlds are q-worlds, then *a fortiori* all p&r-worlds, a subset of the p-worlds, have to be q-worlds. This pattern becomes invalid in the Stalnaker/Lewis analyses. If the p-worlds that are most similar to the evaluation world are all q-worlds, that does not necessitate that the most similar p&r-worlds are also all q-worlds. Lewis (1973) gives a humorous example:

(11) *Failure of Strengthening the Antecedent*
If kangaroos had no tails, they would topple over. $\not\Rightarrow$
If kangaroos had no tails but used crutches, they would topple over.

The variably strict analyses explain why the inference in (11) fails: the worlds where kangaroos have no tails but that are otherwise as similar as possible to the evaluation world are not worlds where kangaroos use crutches, so the first conditional does not connect logically to the second conditional.

Other patterns that are expected to be valid under the strict implication analysis but arguably aren't are Hypothetical Syllogism and Contraposition:

(12) *Failure of the Hypothetical Syllogism (Transitivity)*
If Hoover had been a Communist, he would have been a traitor.
If Hoover had been born in Russia, he would have been a Communist.
$\not\Rightarrow$ If Hoover had been born in Russia, he would have been a traitor.

(13) *Failure of Contraposition*
If it had rained, there wouldn't have been a terrific cloudburst. $\not\Rightarrow$
If there had been a terrific cloudburst, it wouldn't have rained.

[The example in (12) is from Stalnaker 1968, and the example in (13) is a subjunctive version of an example from Adams 1975.]

The non-monotonic analyses predict both of these invalidities correctly.

4 The Dynamic Strict Analysis

Let us return to the apparent failure of Strengthening the Antecedent. Lewis tried to forestall the idea that what is treated as semantic non-monotonicity in his account could actually be explained in a strict implication account by saying that the contextually relevant set of worlds that the conditional quantifies over is easily shifted in a sequence of sentences. He argued that this move would not be able to explain the well-formedness of what became known as Sobel Sequences:

(14) If the USA threw its weapons into the sea tomorrow, there would be war; but if all the nuclear powers threw their weapons into the sea tomorrow, there would be peace.

Lewis deliberately put this example "in the form of a single run-on sentence, with the counterfactuals of different stages conjoined by semicolons and *but*", suggesting that it would be a "defeatist" move to say that in such a tight sequence the context could shift in response to the introduction of a new antecedent clause.

Defeatist or not, based on an observation by Heim (MIT class handout), von Fintel (2001) develops such an account. Heim had noted that Lewis' Sobel Sequence cannot be reversed:

(15) ??If all the nuclear powers threw their weapons into the sea tomorrow, there would be peace; but if the USA threw its weapons into the sea tomorrow, there would be war.

This is unexpected from the point of view of a semantically non-monotonic analysis. In von Fintel's paper, a dynamic strict analysis is developed in which the antecedent has the potential to expand the "modal horizon", the set of contextually relevant possible worlds which the conditional then ranges over. It is shown that if the expansion of the modal horizon is governed by the same similarity ordering used in the Stalnaker/Lewis systems, the analysis replicates the truth-conditions of those systems for isolated or discourse-initial conditionals. The context shifts become only relevant in sequences of conditionals and then create the appearance of semantic non-monotonicity. One crucial argument von Fintel gives for his account is that negative polarity items are licensed in the antecedent of conditionals and that therefore we would prefer a monotonic analysis. It turns out, however, that only a very special notion of monotonicity (dubbed Strawson Downward Entailingness) holds for von Fintel's conditionals: these conditionals are downward monotone in their antecedent only under the assumption that the initial context is such that the modal horizon is already large enough to be unaffected by any of the conditionals in the sequence. This idea is explored for other puzzles for NPI-licensing in von Fintel 1999. The dynamic strict analysis is developed further by Gillies (2007) and critically compared to a pragmatically supplemented non-monotonic analysis by Moss (2010).

5 The Restrictor View

The dominant approach to the semantics of conditionals in linguistics is not so much an alternative to the accounts we have discussed so far, and in particular not to the Stalnaker/Lewis analysis, but a radical rethinking of the compositional structure of conditional sentences. It began with Lewis's 1975 paper on adverbial quantification, which dealt with sentences like

(16) If it is sunny, we always/usually/mostly/rarely/sometimes/never play soccer.

Lewis argued that there was no plausible semantics for the conditional connective that would interact compositionally with the adverbs of quantification to give correct truth-conditions for these sentences. Instead, he argued that the *if*-clause added no conditional meaning of its own to the construction. The idea is that the only "conditional" operator in the structure is the adverb and that *if* merely serves to introduce a restriction to that operator. In other words, where naïvely one would have thought that (16) involved the combination of an adverbial quantificational operator with the conditional expressed by *if*, Lewis argued that there was just one operator and that *if* didn't express any kind of conditional operator of its own.

Lewis himself did not generalize this idea; nowhere else in his writings does he give any indication that *if*'s found elsewhere are to be treated on a par with the *if* in adverbially quantified sentences. (It should be noted that in the adverbial quantification paper, Lewis does suggest that the *if* found in construction with probability operators is also not a conditional operator of its own, although he doesn't say whether it is to be seen as a restrictor in those cases. It is a shame that Lewis did not connect his insights in the adverbial quantification paper to the problems surrounding conditional probability, as discussed for example in Lewis 1976; cf. also Hájek 1993. Kratzer (1986) does make the connection; for some recent discussion see Rothschild 2010 and Egré & Cozic 2011.)

Kratzer took the logical step and argued that Lewis' idea should be applied to all conditional constructions. She put the point very concisely in Kratzer 1986: "The history of the conditional is the story of a syntactic mistake. There is no two-place *if… then* connective in the logical forms for natural languages. *If*-clauses are devices for restricting the domains of various operators. Whenever there is no explicit operator, we have to posit one."

The central idea is that *if* itself does not carry any distinctive conditional meaning, rather it is, so to speak, a helper expression that modifies various quantificational/modal operators. This doesn't just apply when an overt operator combines with an *if*-clause but also when an *if*-clause occurs on its own with no overt operator in sight. In that case, Kratzer suggest, there must a covert, or at least not obviously visible, operator. What one might call bare indicative conditionals either contain a covert epistemic necessity modal or a covert generic frequency operator (≈ *usually/always)*:

(17) If he left at noon, he's home by now. [epistemic necessity]
(18) If he leaves work on time, he has dinner with his family. [generic frequency]

In bare subjunctive conditionals, one should consider the possibility that the modal form *would* is the operator restricted by the *if*-clause, an idea bolstered by the fact that there are *if*-less would-sentences (see Kasper 1992 and Schueler 2008):

(19) I would have beaten Kasparov.

It should be pointed out that while one may have a desire to have a uniform(ish) analysis of indicative and subjunctive conditionals, partially because both are *if … then* constructions, the restrictor analysis opens up a potentially large gap between them. The uniform presence of *if* would be almost entirely beside the point: how big the difference between the two kinds is depends on what, if any, difference there is between the modal operators present in them.

6 The Limit and Uniqueness Assumptions, Conditional Excluded Middle, and *Might*

Lewis and Stalnaker differ in their assumptions about the similarity ordering. Stalnaker assumes that for any (non-contradictory) antecedent and any evaluation world, there will be a unique most similar antecedent world. Lewis neither makes this Uniqueness Assumption (he calls it "Stalnaker's Assumption") nor the weaker Limit Assumption (that for any antecedent and evaluation world, there is a set of most similar antecedent worlds). Informally, here, we have been using the Limit Assumption but not the Uniqueness Assumption when we talk about the most similar or closest antecedent worlds. For discussion of this difference, see Lewis (1973: 19–21) and Stalnaker (1984: Chapter 7, esp. 140–142); Pollock (1976), Herzberger (1979), and Warmbrod (1982) argue for the Limit Assumption as well. More recently, Williams (2010b) and Swanson (2010) have revisited the topic.

In his argument against the Limit Assumption, Lewis adduces cases that seem to show that the closeness to the actual world can get asymptotically closer:

(20) If this one inch line were more than an inch long, ...

For any world in which the line is 1+x inches long, there will be a world where the line is just a little bit shorter but still more than an inch long. So, there will be no world where the line is as close to its actual length as possible while still more than an inch long. Stalnaker argues that in actual use, natural language would not make the fine-grained distinctions needed to threaten the Limit Assumption in such cases.

Cases that throw doubt on the Uniqueness Assumption are less recherché. Quine's example (1950: 14) will do:

(21) If Bizet and Verdi had been compatriots, Bizet would have been Italian.
(22) If Bizet and Verdi had been compatriots, Verdi would have been French.

If the Uniqueness Assumption were correct, exactly one of (21) and (22) would be true. But they both seem dubious.

A virtue of the Uniqueness Assumption is that it validates the principle of Conditional Excluded Middle: either if p, q or if p, not q. Surprisingly, this validity seems to persist even in the problematic Bizet–Verdi case:

(23) Either if Bizet and Verdi had been compatriots, Bizet would have been Italian, or (if Bizet and Verdi had been compatriots) Verdi would have been French.

Stalnaker suggests that (21) and (22) are semantically indeterminate (because it is indeterminate which resolution of the similarity ordering is contextually salient), but

that (23) is true nevertheless because it would be true under any reasonable resolution of the indeterminacy. He proposes implementing this suggestion in a supervaluation framework. Klinedinst (2011) explores this further and shows how it can make good on the suggestion by von Fintel and Iatridou (2002) and Higginbotham (2003) that a CEM-validating semantics for conditionals is behind the intuitive equivalence of pairs of conditionals under quantifiers (contra Leslie 2009; see also Huitink 2010):

(24) Every student would have failed if he had goofed off.
(25) No student would have passed if he had goofed off.

The validity of CEM also is involved in the intuitive relation between *if*-conditionals and *only if*-conditionals:

(26) Only if the Queen had been home, would the flag have flown.
(27) If the flag had flown, the Queen would have been home.

See Barker 1993 and von Fintel 1997 for discussion.

Another consequence of making the Uniqueness Assumption, is that Stalnaker cannot treat *would* and *might* as duals (as proposed by Lewis). Instead, he suggests that *might* is a higher (usually epistemic) operator that takes a whole counterfactual in its scope. And the embedded *would*-counterfactual is of course analyzed as usual. He gives the following paraphrase relation:

(28) If John had been invited, he might have come to the party.
(29) It might be the case that if John had been invited, he would have come to the party.

Consider the following contrast:

(30) It's not the case that John must come to the party. But he might.
(31) #It's not the case that John would have come to the party if he had been invited. But he might have.

(31) should be fine according to Lewis, it's like saying *not all but some*. Stalnaker explains its anomaly this way: in the first sentence the speaker says that he knows that John wouldn't have come, in the second she says that it's compatible with her knowledge that he would have.

Lewis has a purported counterexample to this analysis. Suppose there is in fact no penny in my pocket, although I do not know it since I did not look. "Then 'If I had looked, I might have found a penny' is plainly false." But it is true that it might be, for all I know, that I would have found a penny if I had looked.

Stalnaker contends that under the epistemic reading of *might*, the *might*-conditional is not in fact plainly false. He concedes a non-epistemic reading where the statement is false, but proposes to capture it by relativizing *might* to a situation where the speaker knows all the relevant facts. "This will yield a kind of quasi-epistemic possibility—possibility relative to an idealized state of knowledge." This reading, as he points out, comes out almost identical to Lewis.

The stakes in this debate are increased by the observation that *might*-conditionals are very easy to read as true. If *might* and *would* are duals, then *would*-conditionals are predicted to be very hard to read as true. In fact, Hájek (2009) claims that (almost) all counterfactuals are false because of this. See DeRose (1999) for an earlier discussion and Hawthorne (2005), Williams (2008, 2010a) for more recent relevant work. One might hope that thinking about the contextual dynamics of *might*-conditionals (see Gillies 2007 and Moss 2010) would defuse the argument.

7 Tense and Aspect

We will not be able to discuss the syntax of conditionals in this chapter (cf. Bhatt & Pancheva 2006) but we should take a look at the morphological fine structure of conditionals. It is quite apparent that in English at least, the indicative/subjunctive classification of conditionals is marked by tense and aspect morphology:

(32) If Grijpstra played his drums, de Gier played his flute.
(33) If Grijpstra had played his drums, de Gier would have played his flute.

The earliest works taking the role of tense and aspect in the semantics of conditionals seriously came from Dudman (1983, 1984, 1988, 1989). Work on the interaction of tense and conditionals in philosophical logic includes Nute (1982, 1991); Slote (1978); Thomason and Gupta (1980), and Thomason (1985). A more recent seminal contribution is Iatridou (2000). Since then there has been a proliferation of work on this topic; see Arregui (2005, 2007, 2009); Copley (2006); Ippolito (2007); Kaufmann (2005); Schulz (2008); von Stechow (2007). Here, we can only introduce some basic facts and generalizations.

The central observation is that what is commonly called subjunctive in "subjunctive conditionals" is an additional layer (or two) of past tense morphology, no matter whether the referred to state of affairs is temporally located in the past, present, or future:

(34) a. If Roman comes to the party tomorrow, it will be a grand success.
b. If Roman came to the party tomorrow, it would be a grand success.
c. If Roman had come to the party tomorrow, it would have been a grand success.

(35) a. If Roman is at the post office now, he is missing the meeting, meeting.
b. If Roman were at the post office now, he would be missing the meeting.
c. If Roman had been at the post office now, he would have been missing the meeting.

(36) a. If Roman left before noon, he arrived in time.
b. If Roman had left before noon, he would have arrived in time.

Iatridou (2000) discusses this basic pattern (although she doesn't discuss the two layer pasts in future or present conditionals) and proposes that the additional past does not serve a temporal function. Instead, she argues that the past tense has a schematic semantics that can be applied both temporally and modally: past is an "exclusion feature". When past is used temporally it marks the times talked about as distinct from file now

of the speaker (an additional wrinkle is needed to explain why past means past rather than non-present = past or future). When past is used modally it marks the worlds talked about as distinct from the actual world of the speaker (this does not mean that modally used past is a counterfactuality marker; rather, the intent is to derive something very much like the Stalnaker-analysis of the import of subjunctive marking).

The alternative to Iatridou's account is to try to maintain that the additional pasts in subjunctive conditionals do after all retain their usual temporal meaning. This idea goes back to Dudman (1983, 1984, 1988, 1989) and has been pursued by Ippolito (2003, 2007) and Arregui (2005, 2009), among others. We do not have the space to survey the details of these accounts. Let's rather look at a simplified sketch. Suppose that the extra layer of past tense marks that what the conditional quantifies over is a set of worlds that were accessible from the evaluation world at a past time but may not be anymore. This is typically embedded in a branching futures version of possible worlds semantics. As the time index progresses, more and more open futures are precluded. Imagine that at some point in time, it was an open possibility that Roman would leave before noon, but by the present time it is settled that he did not. Then, assuming that the conditional employs a "historical necessity"-type of accessibility relation, the time index needs to be moved to the past to make sure that the domain of accessible worlds includes at least some worlds where he did leave before noon. Hence, the need for past tense marking on the modal (would = will + PAST) in (36b); the past tense in the antecedent may be a mere agreement phenomenon.

What then about the indicative conditional in (36a)? Clearly, if we assume a historical necessity modal, at the time of utterance it is already settled whether Roman did or did not leave before noon. So, if there need to be at least some antecedent worlds in the domain of the modal, the covert modal in (36a) cannot be a historical necessity modal. Thus, it is not mysterious why (36a) is naturally analyzed as involving a (covert) epistemic necessity modal.

In this story, then, the difference between indicative and subjunctive is twofold: (i) type of accessibility relation/type of modal (epistemic vs. historical), (ii) time index on the modal (present vs. past). An obvious question is whether these differences cross-cut: are there past epistemic conditionals? are there present historic necessity conditionals? The answer to the second question is possibly yes: *If Roman comes to the party tomorrow, it will be a grand success* might arguably be a non-epistemic conditional. The answer to the first question might be expected to be no, since it is well-known that epistemic modals resist embedding under past tense.

One possibly problematic fact for the view just sketched comes from hindsight counterfactuals (Barker 1998, Edgington 2003):

(37) [A randomly tossed coin comes up heads.]
  a. If you had bet on heads, you would have won.
  b. If you bet on heads, you will win.

While (37a) seems acceptable and true after the coin has come up heads, there is no time in the past at which (37b) would have been rational to assert. While that doesn't mean that there wasn't a time at which the indicative conditional was true, it does throw some doubt on the simple idea that the only difference between (37a) and (37b) is the temporal perspective.

Related Topics

3.12 Indicative Conditionals.

References

Adams, Ernest W. 1970. Subjunctive and indicative conditionals. *Foundations of Language* 6(1). 89–94.

Adams, Ernest W. 1975. *The logic of conditional: an application of probability to deductive logic*, Dordrecht: Reidel.

Anderson, Alan Ross. 1951. A note on subjunctive and counterfactual conditionals. *Analysis* 12(2). 35–38.

Arregui, Ana. 2005. *On the accessibility of possible worlds: The role of tense and aspect*. Amherst, MA: University of Massachusetts at Amherst dissertation.

Arregui, Ana. 2007. When aspect matters: The case of *would*-conditionals. *Natural Language Semantics* 15(3). 221–264.

Arregui, Ana. 2009. On similarity in counterfactuals. *Linguistics and Philosophy* 32(3). 245–278.

Barker, Stephen. 1993. Conditional excluded middle, conditional assertion, and *only if*. *Analysis* 53(4). 254–261.

Barker, Stephen J. 1998. Predetermination and tense probabilism. *Analysis* 58(4). 290–290.

Bhatt, Rajesh & Roumyana Pancheva. 2006. Conditionals. In *The Blackwell companion to syntax*, vol. 1, 638–687. Oxford: Blackwell.

Copley, Bridget. 2006. Temporal orientation in conditionals (or, how I learned to stop worrying and love UFOs). Ms, CNRS/Université Paris 8.

DeRose, Keith. 1999. Can it be that it would have been even though it might not have been? *Noûs* 33(sl3). 385–413.

Dudman, V.H. 1983. Tense and time in English verb clusters of the primary pattern. *Australian Journal of Linguistics* 3(1). 25–44.

Dudman, V.H. 1984. Parsing *if*-sentences. *Analysis* 44(4). 145–153.

Dudman, V.H. 1988. Indicative and subjunctive. *Analysis* 48(3). 113–122.

Dudman, V.H. 1989. Vive la révolution! *Mind* 93(392). 591–603.

Edgington, Dorothy. 2003. Counterfactuals and the benefit of hindsight. In Phil Dowe & Paul Noordhof (eds.), *Cause and chance: Causation in an indeterministic world*, 12–27. London: Routledge.

Egré, Paul & Mikaël Cozic. 2011. *If*-clauses and probability operators. *Topoi* 30(1). 17–29.

von Fintel, Kai. 1997. Bare plurals, bare conditionals, and *Only*. *Journal of Semantics* 14(1). 1–56.

von Fintel, Kai. 1998. The presupposition of subjunctive conditionals. In Uli Sauerland & Orin Percus (eds.), *The interpretive tract*, (MIT Working Papers in Linguistics 25), 29–44. Cambridge, MA: MIT Press.

von Fintel, Kai. 1999. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics* 16(2). 97–148.

von Fintel, Kai. 2001. Counterfactuals in a dynamic context. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 123–152. Cambridge, MA: MIT Press.

von Fintel, Kai & Sabine Iatridou. 2002. If and when *If*-clauses can restrict quantifiers. Ms, MIT.

Gillies, Anthony S. 2007. Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3). 329–360.

Goodman, Nelson. 1947. The problem of counterfactual conditionals. *Journal of Philosophy* 44(5). 113–128.

Hájek, Alan. 1993. *Conditional probability*. Princeton, NJ: Princeton University dissertation.

Hájek, Alan. 2009. Most counterfactuals are false. Draft of monograph.

Hawthorne, John. 2005. Chance and counterfactuals. *Philosophy and Phenomenological Research* 70(2). 396–405.

Herzberger, Hans. 1979. Counterfactuals and consistency. *The Journal of Philosophy* 76(2). 83–88.

Higginbotham, James. 2003. Conditionals and compositionality. *Philosophical Perspectives* 17(1). 181–194.

Huitink, Janneke. 2010. Quantified conditionals and compositionality. *Language and Linguistics Compass* 4(1). 42–53.

Iatridou, Sabine. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31(2). 231–270.

Ippolito, Michela. 2003. Presuppositions and implicatures in counterfactuals. *Natural Language Semantics* 11(2). 145–186.

Ippolito, Michela. 2007. Semantic composition and presupposition projection in subjunctive conditionals. *Linguistics and Philosophy* 26(6). 631–672.

Kanazawa, Makoto, Stefan Kaufmann & Stanley Peters. 2005. On the lumping semantics of counterfactuals. *Journal of Semantics* 22(2). 129–151.

Kasper, Walter. 1992. Presuppositions, composition, and simple subjunctives. *Journal of Semantics* 9(4). 307–331.
Kaufmann, Stefan. 2005. Conditional truth and future reference. *Journal of Semantics 22(3). 231–280.*
Klinedinst, Nathan. 2011. Quantified conditionals and conditional excluded middle. *Journal of Semantics* 28(1). 149–170.
Kratzer, Angelika. 1977. What *must* and *can* must and can mean, *Linguistics and Philosophy* 1(3). 337–355.
Kratzer, Angelika. 1979. Conditional necessity and possibility. In Rainer Bäuerle, Ura Egli & Arnim von Stechow (eds.), *Semantics from different points of view*, 117–147. Berlin: Springer.
Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10(2). 201–216.
Kratzer, Angelika. 1986. Conditionals. *Chicago Linguistics Society* 22(2). 1–15.
Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics and Philosophy* 12(5). 607–653.
Kratzer, Angelika. 2005. Constraining premise sets for counterfactuals. *Journal of Semantics* 22(2). 153–158.
Leslie, Sarah-Jane. 2009. *If, unless,* and quantification. In Robert J. Stainton & Christopher Viger (eds.), *Compositionality, context and semantic values: Essays in honour of Ernie Lepore*, 3–30. Berlin: Springer.
Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.
Lewis, David. 1975. Adverbs of quantification. In Edward Keenan (ed.), *Formal semantics of natural language*, 3–15. Cambridge: Cambridge University Press.
Lewis, David. 1976. Probabilities of conditionals and conditional probabilities. *The Philosophical Review* 85(3). 297–315.
Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10(2). 217–234.
Moss, Sarah. 2010. On the pragmatics of counterfactuals. *Noûs*. Published online at http:dx.doi.org/10.1111/j.1468-0068.2010.00795.x/
Nute, Donald. 1982. Tense and conditionals. Tech. rep. Deutsche Forschungsgemeinschaft.
Nute, Donald. 1991. Historical necessity and conditionals. *Noûs* 25(2). 161–175.
Pollock, John. 1976. *Subjunctive reasoning*. Dordrecht: Reidel.
Quine, Willard Van Orman. 1950. *Methods of logic*. New York: Henry Holt and Company.
Quine, Willard Van Orman. 1960. *Word and object*. Cambridge, MA: MIT Press.
Rescher, Nicholas. 1964. *Hypothetical reasoning*. Amsterdam: North-Holland.
Rothschild, Daniel. 2010. Do indicative conditionals express propositions? Forthcoming in *Noûs*.
Schueler, David. 2008. The syntax and semantics of implicit conditionals: Filling in the antecedent. Los Angeles, GA: UCLA dissertation.
Schulz, Katrin. 2008. Non-deictic tenses in conditionals. *Semantics and Linguistic Theory (SALT)* 18.
Slote, Michael A. 1978. Time in counterfactuals. *The Philosophical Review* 87(1). 3–27.
Stalnaker, Robert. 1968. A theory of conditionals. In Nicholas Rescher (ed.), *Studies in logical theory*, 98–112. Oxford: Blackwell.
Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.
von Stechow, Arnim. 2007. Tense and presuppositions in counterfactuals. Handout from a talk in Oslo, October 10, 2007.
Swanson, Eric. 2010. Conditional excluded middle without the limit assumption. Ms, forthcoming in *Philosophy and Phenomenological Research*.
Thomason, Richmond & Anil Gupta. 1980. A theory of conditionals in the context of branching time. *The Philosophical Review* 89(1). 65–90.
Thomason, Richmond H. 1985. Note on tense and subjunctive conditionals. *Philosophy of Science* 52(1). 151–153.
Veltman, Frank. 1976. Prejudices, presuppositions and the theory of conditionals. In Jeroen Groenendijk & Martin Stokhof (eds.), *Amsterdam papers informal grammar, volume I*, 248–281. Amsterdam: Centrale Interfaculteit, University of Amsterdam.
Veltman, Frank. 2005. Making counterfactual assumptions. *Journal of Semantics* 22(2). 159–180.
Warmbrod, Ken. 1982. A defense of the limit assumption. *Philosophical Studies* 42(1). 53–66.
Williams, J. Robert G. 2008. Chances, counterfactuals, and similarity. *Philosophy and Phenomenological Research* 77(2). 385–420.
Williams, J. Robert G. 2010a. Chancy counterfactuals, redux: Response to Dodd. Ms.
Williams, J. Robert G. 2010b. Defending conditional excluded middle. *Noûs* 44(4). 650–688.

3.14
QUESTIONS

Paul Hagstrom

1 Questions, Interrogatives, and Interrogative Acts

What we make of questions depends on our goals. Coming from semantics and logic, we aim to clarify the relationships among questions and between questions and their (actual or potential, complete or partial) answers. Coming from pragmatics, we want to understand the role of questions in structuring discourse, and how they respond to and change the context. Questions also have a role in the process of rational inquiry and theory development.

First, to disentangle the term "question." One can *ask* a question. In so doing, one performs a speech act, an *interrogative act*, of the sort performed when making an assertion or uttering an imperative. An interrogative act is a kind of request, a directive to answer.

Interrogative acts are often performed using a sentence in a particular syntactic form, an *interrogative* form. The interrogative forms in English often involve a change in word order, use of interrogative words such as *if*, *whether*, and *who*, or a particular intonation contour (1). Interrogatives can stand alone as simple sentences (*matrix* interrogatives), or be *embedded* as an *indirect question* (2), just as an indicative sentence can (3).

(1) What did John buy?
(2) I asked Mary [what John bought].
(3) I told Mary [that John bought a laptop].

Here we will concentrate on a more abstract notion of question: the object an interrogative denotes. A question in this sense is something you can have, discuss, wonder about, dismiss, or revisit.

These conceptual divisions parallel those we make for propositions: Interrogative acts and assertions are speech acts, indicatives and interrogatives are sentence forms, and propositions and questions are abstract referents. These notions are dissociable: One can perform an interrogative act with an indicative form, use an interrogative form to make an assertion, or express a question without an interrogative form.

In what follows, we will be surveying some of the basic phenomena, issues, arguments, and conclusions pertaining to questions. However: Our time together is brief, and the field is vast. Other recent overviews are worth consulting, including Ginzburg (1996), Higginbotham (1996), Groenendijk and Stokhof (1997), Harrah (2002), Hagstrom (2003), and Dekker et al. (2007).

2 Relating Questions to Logic and Semantics

The issue of how to relate questions to logic has been a running theme in the literature on questions. Classically, the study of logic deals with inference, deductions, valid argumentation, relating propositions in terms of truth values. A common goal in the investigation of questions is to describe their relationships to propositions and to other questions with a precision and predictive power comparable to that achieved for propositions—but the centrality of truth values in propositional logic (and its extensions) leaves questions without an obvious niche to occupy. Whatever else we know about questions, the intuition is strong that a question differs from a proposition insofar as a question itself can be neither true nor false. How then can we define notions like entailment or equivalence?

Prior and Prior (1955) make some progress toward an "erotetic logic" (a logic of questions), drawing on early proposals from Whately (1854, 1877), but their conclusion is that questions will not submit to a proper treatment within symbolic logic. However, the developments over the subsequent half century entitle us to consider their pessimism premature. As described below, many current approaches succeed in bridging questions and propositions by viewing questions in terms of their answers.

3 Background on Propositions and Possible Worlds

Let us begin by settling on an characterization of a proposition (informally—for discussion of the nuances and issues that arise here, see Chapters 4.1, 5.1, 5.4, 5.5, 5.10). The universe (or "world") is in some state. We may know some aspects of this state, but we can imagine ways in which they might have been different. We will call the set of all states a universe could potentially be in the set of *possible worlds*. Somewhere in there is the *actual world*, but knowing which it is would require being able to fully specify the state of the actual world.

A *proposition* could be true or false: It is true in some possible worlds, false in others. Knowing a proposition to be true narrows the possible locations of the actual world—the actual world must be in that subset of the possible worlds where the proposition is true. Gaining information (coming to know that some proposition is true) shrinks the set of possible worlds containing viable candidates for being the actual world.

We can thus take a proposition to characterize a set of possible worlds (those in which it is true). From the other direction, we can take any set of possible worlds to define a proposition (although perhaps in only an abstract sense—not every set of possible worlds corresponds to a proposition expressible in natural language).

4 Questions, Answers, and Partitions (Hamblin 1958)

Hamblin (1958) proposed three Postulates (4), relating questions, statements, and answers.

(4) **Postulate 1.** An answer to a question is a statement.
Postulate 2. Knowing what counts as an answer is equivalent to knowing the question.
Postulate 3. The possible answers to a question are an exhaustive set of mutually exclusive possibilities.

Hamblin's first Postulate combines the ideas that (a) an answer is a vehicle for information, and (b) the "unit of information" is the proposition. Specifically, the first Postulate excludes sentence fragments from the domain of answers. While it is felicitous in discourse to respond to "Who left?" by saying "Adam," "Adam" is not an answer to the question—it is an elliptical form of the answer "Adam left" and conveys the proposition that Adam left.

The second Postulate encodes the idea that, just as propositions are distinguished by their truth conditions, questions are distinguished by their answerhood conditions. In response to "Who left?" it seems clear that "Adam left" would count as an answer, whereas "Someday the Sun will go out" would not (even if true). A question divides propositions into those that are possible answers and those that are not. Postulate 2 illustrates a specific case, but (even without confronting the issue of what Hamblin meant by "knowing the question") we can take it more generally as setting up an equivalence between a question and its answers.

The third Postulate characterizes the set of answers. A question presents a partition of the possible worlds into distinct cells, such that any given possible world is compatible with exactly one of the answers. To ask a canonical question is to present the hearer with this set of cells and request help locating the actual world in one of them: Which, of these ways that the world could be, is it?

5 The Possible, the True, and the Exhaustive

Before turning to broader empirical questions, we will review a few other analyses for comparison, each of which builds on Hamblin's (1958) insight. One issue of particular concern to these subsequent approaches is the ability to derive appropriate meanings compositionally from the syntactic form (see also Chapters 2.9, 4.6).

5.1 The Tale of Adam, Betty, Christy, and David

Listen:

Once upon a time, there were four students. Their names were Adam, Betty, Christy, and David. Adam and Christy were spies, working for Mr. X. One evening, all four went to a party. But before it was over, Christy and David left.

The End.

5.2 Sets of Possible Answers (Hamblin 1973)

Hamblin (1973) proposed an extension of Montague's (1970) *English as a Formal Language* to capture matrix questions. In so doing, he reinterprets the denotation of a question as being the set of its *possible answers*, abandoning his earlier Postulate 3 (and the partition arising therefrom). The set of possible answers does not partition possible worlds because they can overlap—more than one might be true in a given possible world. The advantage of this view is the ability to derive the set of possible answers from the syntactic form of the question. The question *Who left early?* in the scenario given above has the denotation AnsPoss (5).

(5) AnsPoss: {Adam left early, Betty left early, Christy left early, David left early}

Hamblin (1973) proposed that (a) question words like *who* denote sets of individuals (e.g., *who* denotes {Adam, Betty, Christy, . . .}), and (b) the mechanism of semantic composition operates on sets. A predicate (set) (like {LEFT}) is applied to an argument (set) by "distributing" the application of the predicate(s) "pointwise" over the members of the argument, collecting the results in a set. The derivations of *Adam left* and *Who left?* are sketched below to illustrate.

(6) Adam left.
  a. {LEFT}({ADAM})
  b. {LEFT(ADAM)}

(7) Who left?
  a. {LEFT}({ADAM, BETTY, CHRISTY, DAVID, . . .})
  b. {LEFT(ADAM), LEFT(BETTY), LEFT(CHRISTY), LEFT(DAVID), . . .}

A derivation ending with a singleton set has derived a proposition, and a derivation ending with a multiple-membered set has derived a question. In Hamblin's system, all of the work in question formation is done by the question word: In (7), *who* introduces a multi-membered set into the derivation, and this set-hood "propagates" out to the end of the derivation, yielding a set of propositions.

Note that the ANSPOSS set in (5) does not contain the proposition *nobody left*. Under this analysis, absent some more elaborate story, the response *nobody left* must be treated not as a possible answer, but as a rejection of the question.

Hamblin's (1973) proposed set-based compositional mechanism was not initially widely used, but has more recently been repopularized through Rooth's (1985) influential work on focus semantics, and has since found its way back to analyses of questions as well (see, e.g., Rullmann and Beck 1997; Hagstrom 1998; Kratzer and Shimoyama 2002; Beck 2006; Cable 2007). The term "Hamblin-semantics" in recent literature generally refers to the compositional procedure reviewed above. See also Groenendijk and Roelofsen's (2009) development of a similar analysis ("inquisitive semantics") within a dynamic semantics framework (Chapter 2.7) and with further elaboration concerning interrogative force and disjunctions.

5.3 Sets of True Answers (Karttunen 1977)

Karttunen (1977) updated Hamblin's (1973) proposal for Montague's (1973) *PTQ* framework, aiming to account for embedded questions—which Karttunen took to be primary (reducing matrix questions to embedded questions under an implicit interrogative speech act "I ask you (to tell me)").

The mechanics of Karttunen's proposal need not concern us here, apart from noting that his derivation proceeds by adding a question-forming operator (which we might syntactically identify as a complementizer) to a proposition, followed by composition of any question words. This structure lines up well with the popular hypothesis within generative syntax (Huang 1982) that all question words move to a structurally high position in the syntax prior to interpretation—Karttunen's proposal can be viewed as providing a kind of semantic motivation for this syntactic operation. This synergy may, in part, explain the widespread adoption of Karttunen's semantic analysis—but, whatever the reason, Karttunen's (1977) proposal quickly became the "standard view."

Karttunen's final denotation for questions differs from Hamblin's in one essential respect: While Hamblin's denotation contains the set of *possible* answers, Karttunen's contains the *true* answers. This revision is motivated by the interpretation of questions embedded under verbs like *know* and *tell*. If *know* is given a distributive analysis when provided with a set of propositions (that is, to know a set of propositions is to know each of the propositions in the set), the interpretation of (8) indicates that only the true propositions should be in the denotation of the embedded question. Accordingly, Karttunen's denotation for the question *Who left early?* is AnsTrue (9), "the propositions that jointly constitute a true and complete answer to the question."

(8) Mr. X knows who left early
(9) AnsTrue: {Christy left early, David left early}

As in Hamblin's (1973) approach, *nobody left early* will never be a member of AnsTrue, although Karttunen anticipated the possibility that AnsTrue might be empty in his proposed semantics for *know* (see section 6.8).

5.4 Complete Answers under Partitions (Groenendijk & Stokhof 1982, Higginbotham 1993)

To AnsPoss and AnsTrue above, we can add AnsExh (10), reflecting Hamblin's (1958) partition proposal, where only one member of AnsExh is true in any given possible world.

(10) AnsExh: {Nobody left early, Only Adam left early, Only Betty left early, Only Christy left early, Only David left early, Only Adam and Betty left early, Only Adam and Christy left early, Only Adam and David left early, Only Betty and Christy left early, Only Betty and David left early, Only Christy and David left early, Everyone but Adam left early, Everyone but Betty left early, Everyone but Christy left early, Everyone but David left early, Everyone left early}

More developed partition proposals are outlined by Groenendijk and Stokhof (1982) and Higginbotham (1993). Groenendijk and Stokhof propose that a question is represented as a *contingent proposition*, or a propositional concept. This is a proposition whose extension in any possible world is the true and complete answer in that world. When the propositional concept *whether Adam left early* is evaluated with respect to a world in which Adam left early, it amounts to the proposition *that Adam left early*; when evaluated with respect to a world in which Adam did not leave early, it amounts to the proposition *that Adam did not leave early*. This captures rather neatly the fact that *Mr. X knows whether Adam left early* means that Mr. X knows the true answer to the question.

Under the partition view, *nobody left early* has full citizenship among the possible answers, since it describes one of the cells of the partition of how the world might be.

5.5 Context Dependence in Questions-as-Answers Approaches

In the illustrations of AnsPoss, AnsTrue, and AnsExh above, the answers were limited to propositions concerning the protagonists and situations in the Tale. This type of context-sensitivity is a relatively deep component of questions—were the question

"Who left early?" posed in the context of the Tale, the response "Vyacheslav Molotov" does not seem to constitute an answer (despite being true: in 1945, Vyacheslav Molotov did leave early from a meeting with U.S. President Harry S. Truman).

One way in which context can manifest itself is as a constraint on the domain of individuals *who* quantifies over or refers to—perhaps comparable to contextual constraints on the interpretation of *Everybody left* (which can be evaluated as true even where not all of humanity left; see, e.g., von Fintel 1994). Another way that context can play a role is in ruling out otherwise viable answers; for example, if the context contains a presupposition that exactly one person left early (perhaps contributed by a singular *which* phrase, as in *which person left early?*), then answers entailing that several people left early cease to be viable.

Integrating context sensitivity into an understanding of questions raises several issues. Here are two: If a question is defined in terms of its answers and the answers depend on context, do we conclude that the same interrogative refers to different questions in different contexts? When an otherwise possible answer conflicts with a presupposition in the context, does it cease to be a possible answer for the purposes of defining the question? (Section 6.5 below contains some further discussion of contextual effects.)

5.6 Other Proposals

The analyses outlined above (resulting in AnsPoss, AnsTrue, and AnsExh) are the most influential and commonly discussed, but of course other proposals have been advanced as well. Let me briefly mention two, both opposed to the reduction of questions to answers: Krifka (2001a) argues for a more complex denotation of questions that can apply directly to fragment answers (denying Hamblin's (1958) Postulate 1), and Ginzburg (1995) argues that questions are a basic type and opts for a formalization of discourse rather than properties of questions per se. Neither proposal will be reviewed further here, although we will revisit some of the arguments against reducing questions to answers. Other well-known analyses not even discussed here include Åqvist (1965), Belnap and Steel (1976) and Hintikka (1976).

6 Basic Data and Analytical Issues

6.1 Complete Answers, Partial Answers, and Irrelevant Statements

An advantage that AnsExh ("partition theories") has over both AnsPoss and AnsTrue is that it can distinguish complete answers, partial answers, and irrelevant statements. In (11), AnsExh (10) partitions the set of possible worlds into 16 cells, each corresponding to a possible combination of leavers and nonleavers.

(11) Who left the party early?

a. Christy and David left the party early, and nobody else did.
b. Adam did not leave the party early.
c. Adam brought his camera to the party.

A *complete answer* to the question singles out the cell containing the actual world. Example (11a) is one such proposition; (11b) is not. However, (11b) is a *partial answer*, by virtue of *ruling out* some of the cells—the eight cells in which Adam left early are no

longer viable candidates for containing the actual world. Example (11c), on the other hand, does not appear to rule out any cells, making it an *irrelevant statement*. (Although note: context can affect this. If it is known that, as a rule, when Adam brings his camera to the party, he leaves last—perhaps because his spying duties require him to photograph everyone as they leave—then, we could deduce (11b) from (11c)).

Those who discuss complete answers, partial answers, and irrelevant statements are generally in agreement about the characterizations just given. Although these distinctions are a natural consequence within a partition analysis, it is generally possible to derive a partition even within a nonpartition analysis (see section 6.8).

6.2 Differentiating Yes–No Questions and Alternative Questions

Karttunen (1977) observed an issue with AnsTrue even as he formulated it: The representations of positive and negative yes–no questions are the same. The issue arises equally with AnsPoss and AnsExh. Under all these approaches, a yes–no question divides a proposition p from its negation $\neg p$. Negation in the question will have no effect on this division. Yet, the intuition is that *Did Adam leave early?* and *Didn't Adam leave early?* differ.

Krifka (2001*b*) adds a further observation: Alternative questions where the alternatives are p and $\neg\ p$ also share this division. That is, all three questions in (12) are predicted to have the same denotation. Yet these questions do not behave alike. For example, only (12a) and (12b) admit the responses "yes" and "no"—(12c) requires a more specified response.

(12) a. Did Christy leave?
b. Didn't Christy leave?
c. Did Christy leave or didn't she?

This kind of equivalence is difficult to escape on a partition view. Büring and Gunlogson (2000) and Romero and Han (2004) discuss aspects of the problem, calling upon Ladd's (1981) work on the pragmatic constraints on negative yes–no questions. Both pairs of authors ultimately provide pragmatic proposals, but there is an option for distinguishing these questions within the semantics proper: The answer set for yes–no questions contains only the explicit proposition, while that for alternative questions contains both alternatives ({Christy left} for (12a), {Christy didn't leave} for (12b), {Christy left, Christy didn't leave} for (12c)). Whether such an analysis will ultimately work out remains to be seen (Büring & Gunlogson argue that it won't), but if it does provide a way of extricating ourselves from the equivalence problem, it also positions AnsPoss ahead of the other possibilities for the basic question denotation.

6.3 Semantic Types and Classes of Embedding Verbs

Verbs that permit an interrogative to be embedded as a complement come in at least two varieties. Some, like *wonder* or *ask*, operate on the question itself, as a question, based on the meaning of the verb. Others, like *know* and *tell*, relate an individual and a proposition—even when such verbs take an embedded interrogative, the question is interpreted as its (propositional) answer. A sentence like (13a) puts Mr. X in a "wondering" relation to a question, but (13b) doesn't mean that Mr. X told me the

question—what Mr. X told me was the answer to the question. Verbs that view a question in terms of its answer are also usually able to embed a indicative (14).

(13) a. Mr. X wondered who left the party early.
b. Mr. X told me who left the party early.
(14) Mr. X told me that Christy and David, and nobody else, left the party early.

Proposition-embedding verbs themselves come in at least two classes, those that are *factive*, and those that are not (Kiparsky & Kiparsky 1970). Factive verbs presuppose the truth of their complement. Baker (1968) observed that verbs like *tell*, not normally factive with an indicative complement, become factive with an interrogative complement. That is, in (14), what Mr. X told me could easily be false, but in (13b), what Mr. X told me (the answer to the question) is necessarily true. This effect follows naturally from both the AnsExh and AnsTrue accounts.

One fundamental issue is what place questions (and their answers) should have in our ontology—what are their semantic types? Equally pressing are the issues of what (if anything) differentiates the types of matrix questions and embedded questions, and what (if anything) differentiates embedded questions under *wonder* from those under *know*. Is there a consistent type for questions at all, in either context?

Here are some answers attempted in the literature: Karttunen (1977) takes questions to be sets of propositions, augmenting the semantics of embedding verbs like *know* to convert questions into propositions. Ginzburg (1992) argues that questions should be considered to be an irreducible basic type. Krifka (2001*a*) proposes that the type of a question is not uniform but is determined by the type of a given fragment answer. Groenendijk and Stokhof (1982) take questions and propositions to have the same type under *know*, while questions under *wonder* are propositional concepts.

In support of this last view, Groenendijk and Stokhof (1982) observe that it is possible to conjoin interrogatives and indicatives, as in (15). Given the common assumption that only like categories can be conjoined, this puts embedded interrogatives and embedded indicatives in the same class.

(15) a. John knows that Peter has left for Paris, and also whether Mary has followed him.
b. Alex told Susan that someone was waiting for her, but not who it was.

If questions are taken to have a uniform type, the differentiation between questions under *know* and under *wonder* is handled either by augmenting the lexical semantics of the verbs (adding an operator that converts from the basic type when needed), or by providing the semantic interpretation procedure with a "type-shifting" function that converts one type to another on demand when required. This latter class of analyses includes Lahiri's (2002) "Interrogative Raising" rule, a syntactic rule that moves an embedded question when a type clash necessitates it (much as "Quantifier Raising" moves quantifiers).

A related challenge is the existence and interpretation of *concealed questions* like (16), where what appears to be a noun phrase in the syntax is interpreted as an identity question (17). Concealed questions are discussed by, among others, Baker (1968), Heim (1979), Romero (2005, 2006), Nathan (2006).

(16) Leslie has forgotten the capital of Vermont.
(17) Leslie has forgotten what the capital of Vermont is.

6.4 Exhaustivity and Mention-Some Readings

Groenendijk and Stokhof (1982) explore the properties embedded questions have in logical arguments, and observe that both (18) and (19) are intuitively valid arguments (at least here, where the domain of people is known).

(18) Mr. X knows who left early
Christy left early

Mr. X knows that Christy left early

(19) Mr. X knows who left early
Adam did not leave early

Mr. X knows that Adam did not leave early

Groenendijk and Stokhof observe that only (18) follows from AnsTrue, while both follow from AnsExh (which they advance as an argument for AnsExh). In the context above, AnsTrue is *weakly exhaustive*, allowing the conclusion that Mr. X knows of all those who left early that they left early, while AnsExh is *strongly exhaustive*, allowing the further conclusion that Mr. X also knows of all those who didn't leave early that they didn't leave early.

It is an empirical (and somewhat tricky) question whether our judgments warrant a strongly exhaustive notion of questions in all cases. Berman (1991) claims that it is possible for Mr. X to be *surprised* about who left early without being surprised about who didn't leave early. This judgment pattern is (by design) predicted to be impossible under a strongly exhaustive interpretation like AnsExh—and so, to the extent that such judgments are possible, weakly exhaustive meanings like AnsTrue are still needed (see also section 6.8).

Another challenge for strong exhaustivity arises when question interpretations seem even weaker than weakly exhaustive. Hintikka (1976) observed readings of questions under *know* that have an "existential" or "mention-some" reading. For example, all that is required for (20) to be true is that Adam knows the location of *some* post office—it is not necessary that Adam knows the location of all post offices.

(20) Adam knows where he can find a post office.

6.5 Resolving Questions and Sufficiency of Answers

To the extent that the denotation of a question is identified with its answers, an issue arises concerning the extent to which the "completeness" of an answer varies with context. Ginzburg (1992) observes that the conditions under which a question is "resolved" are affected both by the asker's goals in considering the question, and aspects of the asker's own knowledge state. Boër (1978) and Hintikka (1981) provide related discussion as well.

Two examples of the effects of context are given below, from Ginzburg (1995). The first example relates resolvedness to the questioner's knowledge, and the second relates it to the contextually relevant grain-size.

First: Someone asks a conference organizer "Who has been attending talks?", and receives in reply a complete list of names. The asker might later characterize this as either: (a) having been told who was attending these talks (if the asker knows who the names refer to), or (b) *not* having been told who was attending these talks (if the asker is a reporter, for whom the names are not meaningful). Even a completely exhaustive and true answer may still fail to resolve the question.

Second: A traveler is asked where she is and she responds "Amsterdam." This answer indicates either (a) that she knows where she is (if she's just stepped off an airplane), or (b) that she doesn't know where she is (if she's driving around deep in the city). In the first scenario, "Amsterdam" is sufficiently precise to resolve the question; in the second scenario, it is not.

Aloni (2001) approaches these issues within a partition framework by making use of context-dependent "conceptual cover," representing alternative ways of conceptualizing the individuals in a world. This allows influence from a contextually governed "perspective" while still maintaining a context-invariant semantics for interrogatives. Beck and Sharvit (2002) and Williams (2000) also call upon contextually determined covers to divide questions into subquestions, as part of their analysis of the "quantificational variability effect" (exemplified in sentences like *For the most part, Mr. X knows who left the party early*) discussed by Berman (1991), Lahiri (2002), and Groenendijk and Stokhof (1993).

Regardless of whether these particular applications of covers are exactly right, the notion of a cover seems to have the right character for bringing context-dependency into semantics. Covers have been productively put to use in other areas, such as in the interpretation of plurals (Higginbotham 1981; Gillon 1987; Schwarzschild 1992) (see also Chapter 1.5).

The issues concerning "grain-size" in Ginzburg's (1992) examples may submit to treatments within situation semantics (as, in fact, in Ginzburg's own analysis; see also Kratzer (1989, 2002) and Chapter 5.10).

6.6 Questions with Quantifiers and Quantifying into Questions

Questions featuring a quantifier, such as (21), are often ambiguous, interpreted as one of (21a) (the *individual reading*, answered by, e.g., "Professor Smith"), (21b) (the *list reading*, answered with a list, e.g., "John likes Professor Smith, Pete likes Professor Jones, and Bill likes Professor Nelson"), or (21c) (the *functional reading*, answered with a function, e.g., "his advisor"). The distinction between the individual and list readings feels like a scope distinction, parallel to that in *someone loves everyone*—the intuition being that, in (21a), the question (or question word) takes scope over the quantifier, while in (21b) the quantifier takes scope over the question.

(21) Which professor does everyone like?
   a. Which professor is the x such that, for every person y, y likes x?
   b. For every person y, which professor is the x such that y likes x?
   c. What is the function f such that, for every person y, y likes $f(y)$?

The ambiguity of such questions is limited in several ways. One constraint is syntactic, a consequence of grammatical function. Whereas (21) is ambiguous, (22) is not—it has only an individual reading. In (21), the quantifier is the subject, while in (22), the question word is the subject.

(22) Which professor likes everyone?

Another constraint is on the quantifier itself. Few quantifiers lead to a list reading of this kind. Universal quantifiers like *everyone* do, but *few people* and *nobody* do not. The questions in (23) have no list reading, only individual readings ("caffeine-free diet soda") and functional readings ("their least favorite food"). There is some debate about whether quantifiers like (*at least*) *two*, *many*, and *most* can generate a list reading; Chierchia (1993) claims they cannot, Lahiri (2002) claims that they can. But certainly, monotone decreasing quantifiers like *no* or *few* cannot (Lahiri 2002). However, when this type of question is embedded (24), the restriction on the quantifier type is relaxed (Szabolcsi 1997).

(23) a. What did few people buy?
b. What did nobody buy?
(24) a. John listed what everybody bought.
b. John listed what most people bought.
c. John listed what few people bought.
d. John listed what nobody bought (only what people sold).

Several authors have noted an issue these interpretations raise: It is not obvious what quantifying into a question would mean (see, e.g., Chierchia 1993). Quantification is normally defined over semantic objects that evaluate to a truth value. The interpretation of "everyone left early" is something like 'for every person *x*, the proposition that *x* left early is true'. To the extent that a question has a different semantic type (cf. section 6.3), the quantification rule must be modified if it is to be allowed to take scope over a question.

Several analytical options are open. If questions are themselves propositions (as proposed by Groenendijk & Stokhof 1982), there is no issue. If not, the quantificational mechanism can be augmented (Higginbotham 1993), or one can reanalyze the list reading such that quantification into questions is not required. This latter route has been taken in various ways. Pafel (1999) proposes splitting the semantic components finely enough that a quantifier can take scope above a question word but still below the question-forming operator. Berman (1991) and Chierchia (1991), relying on the fact that functional readings do not require the quantifier to take scope over the question, derive list readings from functional readings (taking a list to be essentially an explicit function). Krifka (2001b) analyzes list readings as conjoined interrogative speech acts, thus not resulting from quantification in the usual sense.

6.7 Knowledge and Inquiry

Arguments between approaches to questions often turn on intuitions about questions embedded under *know*, which raises the possibility that the thorny issues in characterizing "knowledge" could be muddying the waters. Cursory comments follow concerning some recent philosophical work that explicitly addresses semantic issues.

Stanley and Williamson (2001) discuss "knowledge-*that*" and "knowledge-*how*," arguing that the latter should be reduced to the former, building on some of the same models of questions we have seen in this chapter.

Schaffer (2007, 2009) casts a critical eye on semantic proposals that reduce "knowledge-*wh*" to "knowledge-*that*." Schaffer's particular objection is that two distinct

questions might be "convergent" (have same true answer), predicting no difference in meaning when embedded under *know*. Schaffer concludes that the representation of knowledge must include not only propositions but also the questions they answer. However, Aloni and Egré (2010) and Aloni et al. (2009) demonstrate that understanding questions in terms of AnsExh (perhaps with the addition of conceptual covers, section 6.5) prevents convergent questions of this sort from arising. With respect to what constitutes knowledge, Kratzer's (2002) discussion of the distinction between facts and true propositions provides an accessible and recent starting point.

Another area where the semantics and logic of questions intersect with traditional philosophical pursuits is in the examination of the role questions and answers play in scientific inquiry; see discussions by Hintikka (1981), Bromberger (1992), and Collingwood (1940). Here, the focus is on the adequacy of answers in advancing theories, what counts as an explanation, and what types of questions are susceptible to explanation-type answers. This work raises the issue of whether a reduction of questions to AnsPoss, AnsTrue, or AnsExh is compatible with situations where the possible answers, or even the forms they might take, are unknown (cf. Bromberger's (1965) "p-predicament"). In general, such considerations indicate that membership in the set of possible answers must be determined by procedure or rule—that a candidate proposition can be identified as being or not being a possible answer upon encountering it.

6.8 Intertranslatability

Adjudicating between the approaches to questions we've seen is not an easy task; each has its own motivations and potential shortcomings, and compelling arguments can be made that reach seemingly conflicting solutions. One might even conclude that more than one representation is necessary. This section focuses on a proposal outlined by Heim (1994) (see also Lahiri 2002) for deriving AnsExh from AnsTrue.

Karttunen (1977) noted a special case in his construction of AnsTrue: If AnsTrue is empty, knowledge of AnsTrue is to be translated as knowledge that AnsTrue is *empty* (rather than as, say, no relevant knowledge). Heim (1994) observes that generalizing this special case, such that knowledge of AnsTrue always translates to knowledge that AnsTrue is whatever it is, has an effect that is in almost every respect equivalent to knowledge of AnsExh. Hence, AnsExh need not be the basic question denotation, as it can be derived as needed in the context of *know* to capture the strongly exhaustive intuitions (reviewed in section 6.4). In Heim's words: "Maybe one never really bears the *know*-relation to a question unless one also knows that the answer to this question is the answer to it."

Heim proceeds to show that the path between AnsTrue and AnsExh is one-directional—AnsExh can be derived from AnsTrue, but not vice versa. This means that AnsTrue is genuinely richer, but is this additional richness needed? Verbs like *surprise* might constitute such a case (if one can be surprised at who left but not at who didn't leave).

7 Concluding Remark

Primarily, the discussion has outlined the ways in which a question can be viewed in terms of its answers and the ways in which this relation can be made explicit, with a

taste of the advantages and disadvantages of each. This conception remains dominant and common to investigations into questions—though, certainly, arguments have been advanced against reducing questions to their answers.

A great deal more can be said about the topics touched upon here (and about those that haven't been), but this sketch of selected issues relevant to the contemporary study of questions should provide a foothold for approaching the primary literature.

References

Aloni, M. (2001), Quantification under conceptual covers, PhD thesis, University of Amsterdam. Aloni, M. and Egré, P. (2010), 'Alternative questions and knowledge attributions', *The Philosophical Quarterly* **60**(238), 1–27.

Aloni, M., Égré, P. and de Jager, T. (2009), 'Knowing whether A or B', *Synthese* pp. 1–27. Online, available http://www.springerlink.com/content/g672737l15732087/fulltext.pdf.

Baker, C. L. (1968), Indirect questions in English, PhD thesis, University of Illinois.

Beck, S. (2006), 'Intervention effects follow from focus interpretation', *Natural Language Semantics* **14**(1), 1–56.

Beck, S. and Sharvit, Y. (2002), 'Pluralities of questions', *Journal of Semantics* **19**, 105–157.

Belnap, Jr., N. D. and Steel, Jr., T. B. (1976), *The logic of questions and answers*, Yale University Press, New Haven.

Berman, S. (1991), On the semantics and logical form of WH-clauses, PhD thesis, University of Massachusetts, Amherst.

Boër, S. E. (1978), ''Who' and 'whether': Towards a theory of indirect question clauses', *Linguistics and Philosophy* **2**, 307–345.

Bromberger, S. (1965), An approach to explanation, in R. J. Butler, ed., '*Analytical philosophy*', 2nd series, Basil Backwell, Oxford.

Bromberger, S. (1992), *On what we know we don't know: Explanation, theory, linguistics, and how questions shape them*, University of Chicago Press; Center for the Study of Language and Information, Chicago; Stanford.

Büring, D. and Gunlogson, C. (2000), 'Aren't positive and negative polar questions the same?', Ms., UCSC.

Cable, S. (2007), The grammar of Q: Q-particles and the nature of wh-fronting, as revealed by the wh-questions of Tlingit, PhD thesis, Massachusetts Institute of Technology.

Chierchia, G. (1991), Functional WH and weak crossover, in D. Bates, ed., '*Proceedings of the Tenth West Coast Conference on Formal Linguistics*', CSLI, Stanford, pp. 75–90.

Chierchia, G. (1993), 'Questions with quantifiers', *Natural Language Semantics* **1**, 181–234.

Collingwood, R. G. (1940), *An essay on metaphysics*, Clarendon Press, Oxford.

Dekker, P., Aloni, M. and Butler, A. (2007), The semantics and pragmatics of questions, *in* M. Aloni, A. Butler and P. Dekker, eds, 'Questions in dynamic semantics', Elsevier, Amsterdam, pp. 1–40.

Gillon, B. (1987), 'The readings of plural noun phrases in English', *Linguistics and Philosophy* **10**, 199–219.

Ginzburg, J. (1992), Questions, queries, and facts: A semantics and pragmatics for interrogatives, PhD thesis, Stanford University, Stanford, CA.

Ginzburg, J. (1995), 'Resolving questions, I', *Linguistics and Philosophy* **18**, 459–527.

Ginzburg, J. (1996), Interrogatives: Questions, facts and dialogue, in S. Lappin, ed., '*The handbook of contemporary semantic theory*', Blackwell, Oxford, pp. 385–422.

Groenendijk, J. and Roelofsen, F. (2009), Inquisitive semantics and pragmatics, in J. M. Larrazabal and L. Zubeldia, eds, '*Meaning, content and argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics and Rhetoric*', University of the Basque Country Press.

Groenendijk, J. and Stokhof, M. (1982), 'Semantic analysis of wh-complements', *Linguistics and Philosophy* **5**, 175–233.

Groenendijk, J. and Stokhof, M. (1993), Interrogatives and adverbial quantification, in K. Bimbo and A. Mate, eds, 'Proceedings of the Fourth Symposium on Logic and Language', Aron Publishers, Budapest, pp. 1–29.

Groenendijk, J. and Stokhof, M. (1997), Questions, in J. van Benthem and A. ter Meulen, eds, '*Handbook of Logic and Language*', Elsevier/MIT Press, Amsterdam/Cambridge, MA.

Hagstrom, P. (1998), Decomposing questions, PhD thesis, Massachusetts Institute of Technology.

Hagstrom, P. (2003), 'What questions mean', *Glot International* **7**, 188–201.
Hamblin, C. (1958), 'Questions', *Australasian Journal of Philosophy* **36**, 159–168.
Hamblin, C. (1973), 'Questions in Montague English', *Foundations of Language* **10**, 41–53.
Harrah, D. (2002), The logic of questions, in D. Gabbay and F. Guenthner, eds, '*The handbook of philosophical logic*', 2nd edition, Vol. 8, Kluwer Academic Publishers, Dordrecht, pp. 1–60.
Heim, I. (1979), Concealed questions, in R. Bauerle, U. Egli and A. von Stechow, eds, '*Semantics from different points of view*', Springer, Berlin.
Heim, I. (1994), Interrogative semantics and Karttunen's semantics for know, in R. Buchalla and A. Mittwoch, eds, 'Proceedings of IATL 1', Akademon, Jerusalem, pp. 128–144.
Higginbotham, J. (1981), 'Reciprocal interpretation', *Journal of Linguistic Research* **3**, 97–117.
Higginbotham, J. (1993), Interrogatives, in K. L. Hale and S. J. Keyser, eds, '*The view from building 20: Essays in honor of Sylvain Bromberger*', MIT Press, Cambridge, MA, pp. 195–227.
Higginbotham, J. (1996), The semantics of questions, in S. Lappin, ed., '*The handbook of contemporary semantic theory*', Blackwell, Oxford, pp. 361–383.
Hintikka, J. (1976), 'The semantics of questions and the questions of semantics', *Acta Philosophica Fennica* **28**(4).
Hintikka, J. (1981), 'On the logic of an interrogative model of scientific inquiry', *Synthese* **47**(1), 69–83.
Huang, C. T. J. (1982), Logical relations in Chinese and the theory of grammar, PhD thesis, Massachusetts Institute of Technology.
Karttunen, L. (1977), 'Syntax and semantics of questions', *Linguistics and Philosophy* **1**, 3–44.
Kiparsky, P. and Kiparsky, C. (1970), Fact, in M. Bierwisch and K. Heidolph, eds, '*Progress in Linguistics*', Mouton, The Hague, pp. 143–173.
Kratzer, A. (1989), 'An investigation of the lumps of thought', *Linguistics and Philosophy* **12**, 607–653.
Kratzer, A. (2002), 'Facts: Particulars or units of information', *Linguistics and Philosophy* **25**, 655–670.
Kratzer, A. and Shimoyama, J. (2002), 'Indeterminate pronouns: The view from Japanese', Paper presented at the 3rd Tokyo Conference on Psycholinguistics.
Krifka, M. (2001*a*), For a structured account of questions and answers, in C. Féry and W. Sternefeld, eds, '*Audiatur vox sapientiae*', Akademie Verlag, Berlin, pp. 287–319.
Krifka, M. (2001*b*), 'Quantifying into question acts', *Natural Language Semantics* **9**, 1–40.
Ladd, D. R. (1981), A first look at the semantics and pragmatics of negative questions and tag questions, in '*Proceedings of the Chicago Linguistic Society* (CLS)', Vol. 17, pp. 164–171.
Lahiri, U. (2002), *Questions and answers in embedded contexts*, Oxford University Press, Oxford.
Montague, R. (1970), English as a formal language, in B. Visentini et al., eds, '*Linguaggi nella società e nella tecnica*', Edizioni di Comunità, Milan, pp. 189–224.
Montague, R. (1973), The proper treatment of quantification in ordinary English, in J. Hintikka, J. Moravcsik and P. Suppes, eds, '*Approaches to natural language*', Reidel, Dordrecht, pp. 221–242.
Montague, R. (1974), *Formal philosophy: Selected papers of Richard Montague*, Yale University Press, New Haven.
Nathan, L. (2006), On the interpretation of concealed questions, PhD thesis, Massachusetts Institute of Technology.
Pafel, J. (1999), 'Interrogative quantifiers within scope', *Linguistics and Philosophy* **22**, 255–310.
Portner, P. and Partee, B., eds (2002), *Formal semantics: The essential readings*, Blackwell Publishing, Oxford.
Prior, M. and Prior, A. (1955), 'Erotetic logic', *The Philosophical Review* **64**(1), 43–59.
Romero, M. (2005), 'Concealed questions and specificational subjects', *Linguistics and Philosophy* **28**(6), 687–737.
Romero, M. (2006), On concealed questions, in J. Howell and M. Gibson, eds, 'Proceedings of SALT 16', CLC Publications, Ithaca, NY.
Romero, M. and Han, C. (2004), 'On negative yes/no questions', *Linguistics and Philosophy* **27**, 609–658.
Rooth, M. (1985), *Association with focus*, PhD thesis, University of Massachusetts, Amherst.
Rullmann, H. and Beck, S. (1997), Reconstruction and the interpretation of which-phrases, in G. Katz, S.-S. Kim and H. Winhart, eds, 'Reconstruction: Proceedings of the 1997 Tübingen Workshop', Vol. 127, Arbeitspapiere des Sonderforschungbereichs 340, pp. 223–256.
Schaffer, J. (2007), 'Knowing the answer', *Philosophy and Phenomenological Research* **75**(2), 383–403.
Schaffer, J. (2009), 'Knowing the answer redux: Replies to Brogaard and Kallestrup', *Philosophy and Phenomenological Research* **78**(2), 477–500.
Schwarzschild, R. (1992), 'Types of plural individuals', *Linguistics and Philosophy* **15**, 641–675.

Stanley, J. and Williamson, T. (2001), 'Knowing how', *Journal of Philosophy* **98**(8), 411–444.
Szabolcsi, A. (1997), Quantifiers in pair-list readings, in A. Szabolcsi, ed., 'Ways of scope taking', Kluwer, Dordrecht, pp. 311–347.
von Fintel, K. (1994), Restrictions on quantifier domains, PhD thesis, University of Massachusetts, Amherst.
Whately, R. (1854), *Elements of logic*, Morton & Griswold, Louisville, KY.
Whately, R. (1877), *Elements of rhetoric: Comprising an analysis of the laws of moral evidence and of persuasion with rules for argumentative composition and elocution*, Harper & Brothers, New York.
Williams, A. (2000), Adverbial quantification over (interrogative) complements, in R. Billerey and B. D. Lillehaugen, eds, '*Proceedings of the 19th West Coast Conference on Formal Linguistics*', Cascadilla Press, Somerville, MA, pp. 547–587.
Åqvist, L. (1965), *A new approach to the logical theory of interrogatives*, Filosofiska Föreningen, Uppsala, Sweden.

Section IV

METHODOLOGY

4.1
THE ROLE OF EXPERIMENT

Edouard Machery and Stephen Stich

In this chapter, we argue that philosophers of language should emulate linguists, who are increasingly replacing the traditional informal reliance on their own and their colleagues' intuitions with systematic experimental surveys of ordinary speakers' intuitions. This experimental turn has enabled linguists to alleviate some significant concerns about their methodology, and a similar experimental turn in the philosophy of language would, we believe, enable philosophers to alleviate similar concerns about *their* methodology.

Here is how we will proceed. In Section 1, we will describe the experimental turn in generative linguistics, and argue that a similar methodological shift is called for in the philosophy of language. In Section 2, we review some of the burgeoning experimental work in the philosophy of language, which illustrates both the need for an experimental approach in the philosophy of language and the benefits to be gained. In the remainder of the chapter, we focus on the most well-known experimental work in the philosophy of language: the cross-cultural studies of intuitions about the reference of proper names. In Section 3, we address an influential objection against this work. In Section 4, we argue that an experimental turn in the philosophy of language is bound to have some significant philosophical implications, illustrating this claim with the implications of the cross-cultural work on intuitions about reference.

1 Emulating the Experimental Turn in Generative Linguistics

1.1 The Experimental Turn in Generative Linguistics

Following Chomsky (1957), most syntacticians in the generative tradition rely on a familiar methodology: develop theories that accommodate judgments about linguistically significant properties of sentences and their parts. While various sorts of judgments have been used, judgments about the acceptability of sentences have played a central role in generative syntax. Because the judgments that serve as evidence for syntactic theories are typically made quickly without being consciously inferred (they spring to mind, as it were), they are often called "intuitions." Syntactic theories in generative

linguistics are evaluated on the basis of their consistency with these intuitions: they are supported if they entail that an intuitively acceptable construction is grammatical or that an intuitively unacceptable construction is not grammatical, and they are undermined if they entail that an intuitively unacceptable construction is grammatical or that an intuitively acceptable construction is not grammatical. Theories are also challenged if they entail that some sentence can be read in two different ways while it is intuitively judged to have a single reading, and in a variety of other ways. Naturally, just as scientists sometimes discard some empirical findings, generative syntacticians also recognize that inconsistencies between a theory and some judgments can be explained away by assuming that for various reasons the judgments are erroneous; these discounted judgments are often called "performance errors." For instance, one might fail to notice one of the readings of a genuinely ambiguous sentence or one might find a grammatical construction (for example, a garden-path sentence) unacceptable because one fails to parse it in the right way.

Until the 1990s, linguists studying a particular syntactic phenomenon typically relied on their own intuitions and on the intuitions of other linguists. For instance, in *The Logical Structure of Linguistic Theory*, Chomsky argues that in English the position of verb particles is determined not by the length of the sentential constituents, but by their complexity. He writes (1975: 477):

> While . . . both "the detective brought in the suspect" and "the detective brought the suspect in" are grammatical, in general the separability of the preposition is determined by the complexity of the NP object. Thus we could scarcely have "the detective brought the man who was accused of having stolen the automobile in." It is interesting to note that it is apparently not the length in words of the object that determines the naturalness of the transformation, but, rather, in some sense, its complexity. Thus "they brought all the leaders of the riot in" seems more natural than "they brought the man I saw in." The latter, though shorter, is more complex.

For present purposes, the important point is that Chomsky supports his hypothesis by appealing to his own intuitions, assuming that these are representative of the intuitions of lay competent speakers of English. (For empirical evidence that this assumption is not entirely correct, see Wasow & Arnold 2005).

In recent years, this feature of the methodology of generative linguistics has been extensively criticized. Schütze, in his landmark study of the methods of generative linguistics, *The Empirical Base of Linguistics* (1996), made a powerful case that relying on syntacticians' intuitions is methodologically inappropriate. Although eliciting the intuitions of ordinary competent speakers can be fraught with difficulties (e.g., Schütze 2005; Fanselow 2007), experimental work on linguists' and ordinary competent speakers' intuitions has not shown that the former are more reliable than the latter, provided that ordinary speakers understand the nature of the task (Schütze 1996; Culbertson & Gross 2009). Thus, syntacticians' reliance on their own intuitions and those of their colleagues cannot be justified on the grounds that their expertise make these intuitions more reliable. At best, this long-standing feature of their methodology could be justified on pragmatic grounds since it is easier for syntacticians to consult their own and their colleagues' intuitions than to survey the intuitions of ordinary competent speakers. This would be entirely reasonable if syntactians' reliance on

their own and on their colleagues' intuitions did not have important drawbacks. But that is not the case. First, and most obviously, syntacticians' theoretical commitments risk influencing their intuitions, undermining the evidential role of these intuitions. Several examples discussed by Schütze show that this worry is not merely speculative (see also Labov 1975). Lasnik and Saito (1984) assert that sentences such as "Why do you think that he left?" are ambiguous; Aoun et al. (1987) claim that they are not. Unsurprisingly, the ambiguity of this type of sentence follows from Lasnik and Saito's theory, but not from Aoun et al.'s theory! Second, and equally important, syntacticians' appeal to their own and to their colleagues' intuitions can blind them to the dialectical variation within a given language and biases their understanding of the variation in syntactic properties. Thus, for example, some linguists (for example, Chomsky 1976) have held that the contraction "wanna" is unacceptable in sentences such as (1) because "who" leaves an unpronounced trace after "want" when it is fronted from its position in (2):

(1) Who did you wanna meet your parents?
(2) You wanted who to meet your parents?

However, some ordinary speakers find sentences such as (1) perfectly acceptable, suggesting that in this respect the grammar of English may vary across individuals (Labov 1996; Wasow & Arnold 2005; but see Phillips 2009 for discussion). Worse, syntactians' intuitions might be entirely unrepresentative of the language they speak. For instance, a number of claims made about idioms have been shown to be inaccurate (Nunberg et al. 1994; Riehemann 2001; for a review of the intricate debate about the "wanna" contraction, see Goodall 2006; see Featherston 2007 and Riemer 2009 for other examples). Though Jackendoff (1997) claimed that the idiom "raise hell" is syntactically inflexible and cannot be turned into a passive construction, Riehemann (2001) showed that passive versions of "raise hell" are actually often used (discussed in Wasow and Arnold 2005: 1487–8). Finally, Wasow and Arnold (2005: 1485–6; see also Sampson 2005) suggest that syntacticians' reliance on their own and their colleagues' intuitions prevents them from paying attention to easily available counterevidence, perhaps because of a confirmation bias of some sort.

Linguists and philosophers have not only criticized syntacticians' reliance on their own and their colleagues' intuitions, they also have put forward an alternative methodology: the careful survey of the intuitions of ordinary competent speakers (Bard et al. 1996; Schütze 1996; Cowart 1997; Marantz 2005; Sorace & Keller 2005; Featherson 2007). Such intuitions are not contaminated by linguistic theories. By surveying them linguistics avoid relying on intuitions which may be unrepresentative and they are more likely to detect dialectical variation within a linguistic community. Furthermore, as even critics of this alternative methodology have acknowledged (Phillips 2009), surveying a large number of ordinary speakers has a number of methodological advantages (for discussion of some of these, see Featherston 2007; Myers 2009). For instance, aggregating a large number of acceptability ratings is likely to be the only reliable way to identify subtle differences in acceptability (Bard et al. 1996; Sorace & Keller 2005; Sprouse 2007).

For these and other reasons, Schütze issues the following recommendation (1996: 212):

> Linguistics has much to gain and nothing to lose by taking data collection, and particularly judgment collection, much more seriously, both with regard to the insights that will be gained and the theoretical issues that will be clarified, and with regard to the standing of the field as a scientific endeavor in the larger academic setting. The realization seems to be growing that the psychology of grammaticality judgments can no longer be ignored.

This methodological recommendation has been positively received among generative syntacticians, who rely increasingly on experimental surveys of people's intuitions (Schütze 2005; Myers 2009). Semanticists also have started using experimental methods (Schütze 2005; for an example, see Pietroski et al. 2009), while the field of experimental pragmatics is burgeoning (Noveck & Reboul 2008).

Work on Picture Noun Phrases (PNPs) provides a good illustration (Keller & Asudeh 2001; see also Featherson (2005) and Haider (2007) on *that*-traces in German). While binding theory originally predicted that anaphors can be bound when pronouns can't, and vice versa (Chomsky 1981), PNPs such as (3) and (4) appear to show that this is not the case:

(3) Hanna$_i$ found a picture of her$_i$.
(4) Hanna$_i$ found a picture of herself$_i$.

To deal with this type of PNP, various modifications of binding theory have been proposed, including the hypothesis that coreference in some constructions (particularly, some PNPs) is governed by pragmatic considerations rather than by syntactic or structural features. Attempting to decide between the proposed hypotheses, Keller and Asudeh (2001) found that the acceptability judgments made by lay speakers of English were at odds with *all* the approaches to coreference in PNPs. Contrary to the contention of pragmatic theories, coreference in PNPs turns out to be influenced by structural features, but in an entirely unexpected manner. Keller and Asudeh conclude (2001: 488) that "contrary to all that has been written in the syntactic literature, anaphors can be bound by the subject even in PNPs with possessors."

1.2 Moral for Philosophers of Language

The similarities between the typical methods in the philosophy of language and in generative linguistics are striking. In both fields, intuitions are an important source of evidence, and philosophers overwhelmingly rely on their own intuitions and those of their colleagues much as generative syntacticians did until the 1990s. We believe that the three reasons we have discussed that have led a growing number of syntactians to replace reliance on their own and their colleagues' intuitions with carefully designed experimental surveys of the intuitions of ordinary competent speakers carry over to the philosophy of language straightforwardly.

(1) Despite pronouncements about the superiority of the intuitions of philosophers of language over those of ordinary people (Ludwig 2007; Devitt 2011), there is in fact no empirical evidence supporting this claim.
(2) Philosophers' theoretical commitments risk influencing their intuitions.
(3) Philosophers' reliance on their own intuitions risk blinding them to the

> diversity of the relevant intuitions (see next section for some robust evidence supporting this claim) or to the fact that their own intuitions are unrepresentative.
>
> (Stich 2010)

Thus there is a strong case to be made for philosophers of language to follow the lead of generative syntacticians and to replace the appeal to their own and their colleagues' intuitions with systematic empirical studies of ordinary speakers' intuitions. In the next section, we review some work that has done just that.

2 Empirical Research in the Philosophy of Language

In this section, we review the burgeoning experimental work in the philosophy of language. We believe that this work vindicates our concerns about traditionalists' methods in the philosophy of language, and illustrates the need for an experimental turn in this field too.

2.1 Cross-Cultural Variation in Intuitions about Reference

As noted in Section 1, syntacticians are replacing the reliance on their own and on their colleagues' intuitions with systematic surveys of the intuitions of competent speakers in part on the grounds that, by relying on the intuitions of a small segment of humanity, they might be led to ignore the diversity of people's intuitions and what that diversity reveals about the language they speak. As we now show, much the same concern applies to the philosophy of language.

Philosophers of language have long been interested in explaining how words (proper names, natural kind terms, etc.) refer, and existing theories of reference can be usefully classified into two distinct types: descriptivist theories and causal-historical theories. Focusing on proper names, descriptivist theories propose that a name refers to whatever individual best satisfies the description competent speakers associate with this name (Searle 1958; Jackson 1998). Causal-historical theories propose that a proper name refers to the individual the proper name was introduced to refer to, provided that there is an appropriate causal chain connecting the introduction of the term with current users of the term. This is the case, causal-historical theories maintain, whether or not the individual satisfies the description competent speakers associate with the name (Kripke 1972/1980; Devitt 1981).

To decide between competing theories of reference, philosophers of language have relied on what Machery et al.(2004) have called the method of cases (for discussion, see Deutsch 2010 and Devitt 2011). The method of cases appeals to the fact that people—philosophers, linguists, and ordinary speakers alike—appear to have intuitions about what words refer to in actual and possible situations. Appealing to these intuitions, the method of cases proposes that the correct theory of reference for words of type T is the one which is best supported by competent speakers' intuitions about the reference of words of type T in actual and possible situations.

To see more concretely how the method of cases works, we briefly review one of Kripke's most widely discussed cases. In this case ("the Gödel case"), a name is widely associated with a description that is false of the original bearer of that name, *a*, but true of some other person, *b*. Because descriptivist theories of reference hold that a name

refers to the individual that best satisfies the description competent speakers associate with it, descriptivist theories entail that the name in Kripke's example refers to *b*, the satisfier of the description. But, Kripke maintains, this is just wrong.

> Suppose that Gödel was not in fact the author of [Gödel's] theorem. A man called "Schmidt," whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel.

Now Kripke asks us to consider an "ordinary man" who is well educated in mathematics and can give an account of the incompleteness theorem. But, like many other people, practically the only thing he has heard about Gödel is that he discovered the incompleteness of arithmetic.

> On the [descriptivist] view . . . when our ordinary man uses the name "Gödel," he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description "the man who discovered the incompleteness of arithmetic." . . . But it seems we are not. We simply are not.
>
> (Kripke 1972/1980: 83–4)

In contrast, causal-historical theories of reference are consistent with the intuition that "Gödel" continues to refer to its original bearer, *a*, because *a* is the person causally-historically linked with contemporary uses of the name. It is clear that most philosophers share Kripke's intuition about this case (and other cases), and even contemporary descriptivists acknowledge that these intuitions have falsified traditional forms of descriptivism (Evans 1973; Jackson 1998).

But how widely shared are these intuitions by competent speakers who are not professional philosophers? On the basis of Nisbett and colleagues' findings about the variation of cognitive styles across cultures (Nisbett 2003), Machery and colleagues (2004) hypothesized that the intuition elicited by the Gödel case might differ between Eastern and Western cultures. To test this hypothesis, they constructed two vignettes, "the Gödel vignette" and "the Tsu Ch'ung Chih vignette." The Gödel vignette closely followed Kripke's own wording of the case:

> Suppose that John has learned in college that Gödel is the man who proved an important mathematical theorem, called the incompleteness of arithmetic. John is quite good at mathematics and he can give an accurate statement of the incompleteness theorem, which he attributes to Gödel as the discoverer. But this is the only thing that he has heard about Gödel. Now suppose that Gödel was not the author of this theorem. A man called "Schmidt," whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and claimed credit for the work, which was thereafter attributed to Gödel. Thus, he has been known as the man who proved the incompleteness of arithmetic. Most people who have heard the name "Gödel" are like John; the claim that Gödel discovered the incompleteness theorem is the only thing they have ever heard about Gödel. When John uses the name "Gödel," is he talking about:

(A) the person who really discovered the incompleteness of arithmetic?
or
(B) the person who got hold of the manuscript and claimed credit for the work?

The Tsu Ch'ung Chih vignette had the same structure, but it used names of Chinese individuals (Machery et al., 2004: B9–10). Participants were undergraduates at Rutgers (U.S.A.) and at the University of Hong-Kong, an English-speaking university (for more detail, see Machery et al. 2004).

As Machery and colleagues had predicted, Chinese participants were significantly more likely than American participants to have descriptivist intuitions—that is, intuitions in line with what descriptivist theories of reference say the relevant proper names refer to (see Figure 4.1.1). In fact, most Chinese turned out to have descriptivist intuitions, while most Americans turned out to have causal-historical intuitions—intuitions in line with what causal-historical theories of reference say the relevant proper names refer to. Machery and colleagues also found a surprising amount of within-culture variation. For instance, while a majority of American participants had causal-historical intuitions, about 40 percent had descriptivist intuitions. Machery et al. (forthcoming) have replicated these findings, while Machery et al. (2010) have shown that Chinese participants have similar intuitions when the vignette is written in English and when it is written in Cantonese.

2.2 Intuitions about Truth-Value

Following Kripke's own wording of the Gödel case and the standard practice among philosophers working on reference (for example., Evans 1973), Machery et al. (2004) examined people's intuitions about the reference of a proper name—namely intuitions expressed by sentences such as "When *x* uses the name *y*, he is talking about . . ." or "*y* refers to . . .". Other intuitions, such as intuitions about the truth-value of sentences containing names, are also relevant for theories of reference (for discussion, see, for

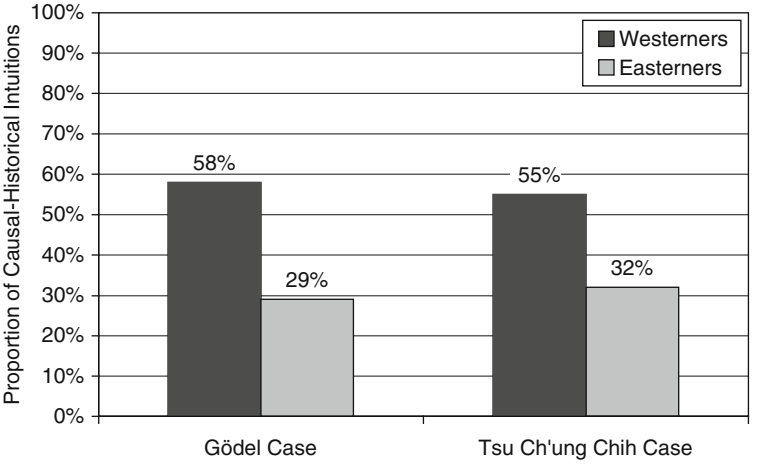


Figure 4.1.1 Results from Machery et al. (2004)

example, Martí 2009 and Machery 2011). Machery et al. (2009) examined whether intuitions about the truth-value of sentences containing names and intuitions about the reference of these names diverge or, rather, whether they are congruent. They presented participants from three countries (Mongolia, India, France) with one of the following two vignettes, and Machery and Olivola have recently done the same thing with American participants (unpublished data):

> Ivy is a high school student in Hong Kong. In her astronomy class, she was taught that Tsu Ch'ung Chih was the man who first determined the precise time of the summer and winter solstices. But, like all her classmates, this is the only thing she has heard about Tsu Ch'ung Chih. Now suppose that Tsu Ch'ung Chih did not really make this discovery. He stole it from an astronomer who died soon after making the discovery. But the theft remained entirely undetected and Tsu Ch'ung Chih became famous for the discovery of the precise times of the solstices. Everybody is like Ivy in this respect; the claim that Tsu Ch'ung Chih determined the solstice times is the only thing people have heard about him. Having read the above story and accepting that it is true, when Ivy says, "Tsu Ch'ung Chih was a great astronomer," do you think that her claim is: (A) true or (B) false?

The second vignette was identical except for the question, which was:

> Having read the above story and accepting that it is true, when Ivy uses the name "Tsu Ch'ung Chih," who do you think she is actually talking about:
>
> (A) the person who (unbeknownst to Ivy) really determined the solstice times?
>
> or
>
> (B) the person who is widely believed to have discovered the solstice times, but actually stole this discovery and claimed credit for it?

Both vignettes are inspired by Kripke's Gödel case, and they are similar to the Tsu Ch'ung Chih vignette used in Machery et al. (2004).

Two findings emerge from this study (see Figure 4.1.2). First, in four different countries, which vary tremendously in terms of culture, the proportion of causal-historical intuitions is similar when participants are asked about the truth-value of a sentence containing a proper name or about the reference of this name, suggesting that people's intuitions about the truth-value of sentences containing proper names track their intuitions about the reference of proper names. When someone judges that a proper name *x* refers to a particular individual *y* in a fictional case, her intuition about the truth of a sentence containing *x* in this case seems to depend on what is said about *y*. Second, the study found further cross-cultural variation in the intuitions that are relevant for studying reference. Consistent with previous work, Americans are more likely to have causal-historical intuitions about the reference of proper names than people from Mongolia, India, and France when asked about what a proper name refers to, and they were more likely than French people to have causal-historical intuitions when asked about the truth-value of a sentence containing proper names.

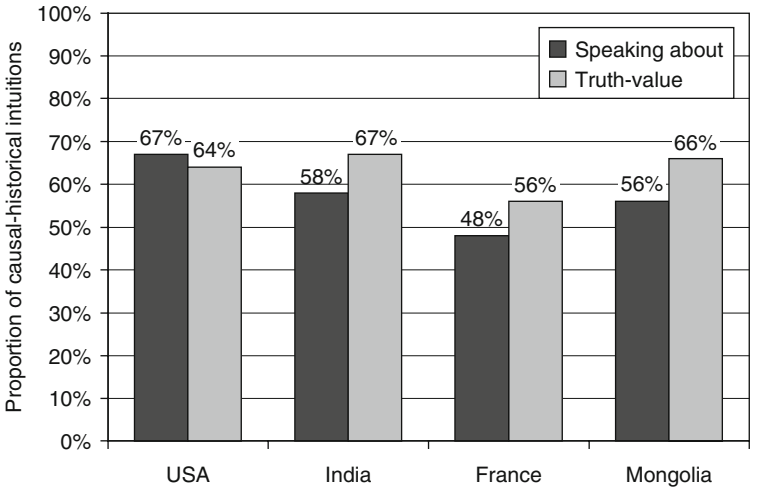


Figure 4.1.2 Results from Machery et al. (2009)

2.3 Externalist Intuitions

Empirical work on intuitions relevant to the philosophy of language is not restricted to intuitions about proper names. Jylkkä et al. (2009) have examined whether lay speakers have externalist intuitions (see also Genone and Lombrozo forthcoming; for relevant work in psychology, see Braisby et al.1996; Hampton et al. 2007). Externalism about natural kind terms is, roughly, the view that the reference of natural kind terms (for example, "gold" and "cat") is determined by facts about substances and kinds about which speakers might have no knowledge (such as the essence of chemical kinds), while internalism is, roughly, the view that the reference of natural kind terms is determined by the descriptions speakers associate with these terms (Putnam 1975; for review, see Wikforss 2008). Furthermore, some externalists hold that externalism is true in virtue of the way competent speakers use natural kind terms: A natural kind term *t* refers in an externalist manner if and only if competent speakers intend *t* to refer to a class of entities that share some fundamental properties that may be unknown to the speakers (for example, a particular genetic structure for the entities in the extension of "cat").

To examine whether people have externalist or internalist intuitions, Finnish participants were presented with vignettes like the following:

> A yellowish, bitter-smelling, fragile mineral called zircaum occurs widely in Mid-Siberian soil. Scientists generally believe that zircaum is the compound ACB. In Northern Norway, a deposit of a substance just like zircaum is found—it is yellowish, bitter-smelling, fragile, etc. When scientists examine its deep structure, they conclude that it is ACB, just like zircaum is believed to be.

At this point, participants were asked whether the novel substance is zircaum on a 7-point scale (question 1). Then, participants were presented with the following passage:

> A few weeks after the discovery in Northern Norway, scientists examine the Mid-Siberian substance more closely. Using methods and instruments more

> exact than previously available, they find out that they were wrong about the deep structure of the substance: the substance is KML instead of ACB. However, the substance found in Northern Norway was indeed ACB, just as the scientists thought it was.

Participants were then asked whether they considered their earlier judgment justified (question 2) and correct (question 3):

> When in the earlier situation you judged that 'the substance found in Northern Norway is/is not zircaum', was your answer a) justified b) strictly speaking correct?

Jylkkä and colleagues found that on average Finnish participants harbored externalist intuitions (a second experiment replicated these findings). Concerning question 1, when a novel substance had the chemical structure of a known element (for example, zircaum), participants tended to judge that it was a sample of this element, while when the novel substance did not have the chemical structure of a known element, they tended to judge that it was not a sample of this element (*mutatis mutandis* for species). Concerning question 3, two thirds of participants said that their answer to question 1 was mistaken after having been told that the beliefs about the constitution of the known substance (for example, zircaum in the story above) were mistaken. Thus, most Finnish participants reported externalist intuitions. However, consistent with our hypothesis that philosophers of language might be led to ignore the variation in intuitions when they appeal only to their own intuitions, Jylkkä and colleagues also found that intuitions about natural kind terms vary substantially among Finnish people. Indeed, Jylkkä and colleagues also report that people were rarely consistently externalist or internalist.

2.4 Conclusion

The findings reviewed in Section 2 provide support for our claim that philosophers of language would be wise to emulate the experimental turn in generative linguistics since some of the intuitions that are of most interest to philosophers of language appear indeed to vary across and within cultures. Philosophers of language exclusive reliance on their own and their colleagues' intuitions obscures much of this variation and makes it more likely that they will develop and embrace philosophical theories that are unable to account for the variation of these intuitions across and within cultures.

3 Responding to Some Criticisms of Recent Empirical Research in the Philosophy of Language

In the remainder of this chapter, we focus on the most developed body of experimental evidence bearing on the philosophy of language—namely the cross-cultural research on intuitions about the reference of proper names. The findings reviewed above have been criticized on various grounds (Ludwig 2007; Martí 2009; Deutsch 2009, 2010; Jackman 2009; Lam 2010; Devitt 2011; Sytsma & Livengood 2011 Ichikawa et al., forthcoming). We will focus on the most frequently raised criticism (Ludwig 2007; Deutsch 2009). Because the question participants were asked in Machery et al. (2004) failed to distinguish between speaker's reference and semantic reference, the critics argue, it is unclear

whether the cross-cultural data bear on theories of reference and should be of interest to philosophers of language. In this section, we first spell out the objection in greater detail, and then offer two distinct replies.

3.1 Speaker's Reference and Semantic Reference

The vignettes used in Machery et al. (2004) and in several other articles ask participants whom a speaker is talking about when he uses a proper name ("Gödel" or "Tsu Ch'ung Chih"). Ludwig (2007) and Deutsch (2009) have complained that this question is ambiguous since it could be interpreted as asking either (i) What is the speaker's reference of a given token of "Gödel"? or (ii) What is the semantic reference of "Gödel"? The semantic reference of a proper name is the reference it has in virtue of the conventional rules of the natural language this name belongs to (Kripke 1977). Thus, "Barack Hussein Obama" refers to Barack Obama in virtue of the rules of English. But in some contexts a speaker may use a given proper name to refer to an individual who is not the semantic reference of this name. For instance, suppose that in 2006 John sees on TV an individual in Baghdad who looks like Dick Cheney and that John says, "Cheney is wearing a nice suit." In reality, Cheney, who is hiding in his ultra-secret bunker, has sent a look-alike to Baghdad. Because John intends to refer to the individual he sees on TV, the speaker's reference of his utterance of "Cheney" is the look-alike in Baghdad, even though the semantic reference of "Cheney" is the former vice-president. Furthermore, Ludwig and Deutsch hold that only intuitions about semantic reference are relevant in determining how the reference of proper names is fixed. While philosophers of language are likely to understand that the question in the Gödel vignette is meant to bear only on semantic reference, lay competent speakers might not be aware of the distinction between these two ways of referring or, if they are aware of this distinction, they might not realize that the question is meant to elicit intuitions about semantic reference. It is thus unclear whether their answers express a judgment about the semantic reference of "Gödel" or "Tsu Ch'ung Chih" in the vignettes used in past research. Deutsch and Ludwig conclude that the experiments conducted by Machery and colleagues fail to provide any evidence that the relevant intuitions about reference vary across cultures.

3.2 Is This A Plausible Objection?

Let's begin with a few uncontroversial points. It makes sense to ask about speaker's reference only for the occurrences of a proper name, and not for a proper name *qua* type. In addition, to identify the speaker's reference of the occurrence of a proper name, one needs to know the communicative intention of the speaker. Thus, in the example given above, we can determine that the speaker's reference of "Cheney" is the look-alike in Baghdad because the cover story and the sentence in which the proper name "Cheney" is embedded provide enough information about John's communicative intention: He intends to talk about the suit of the individual he is seeing on TV. If no information were presented—for instance, if we were just asked who "Cheney" refers to when John uses it—we would focus on the semantic reference of "Cheney" because the question would seem to bear on the reference of this proper name *qua* type and because no contextual information would be provided that would allow us to identify the speaker's communicative intention.

Now consider the Gödel vignette in Machery et al. (2004). At the end of the vignette, participants are asked who John is talking about "when he uses the name 'Gödel'." Since no specific utterance is mentioned and no contextual information is provided that would enable participants to determine John's communicative intention, it is hard to see how participants could understand the question to be about the speaker's reference of John's utterance of "Gödel." Rather, we submit, on the only plausible interpretation of the question, it is asking about the reference of "Gödel" *qua* type—that is, it is asking about the semantic reference of the term in John's language. If this is right, then, *pace* Ludwig and Deutsch, the question in the probes is *not* ambiguous.

3.3 Some Preliminary Empirical Evidence

Though we doubt that the question posed to participants in the Machery et al. (2004) study is ambiguous, others may not be convinced. This motivated Machery et al. (forthcoming) to design a further study in which they attempt to eliminate the ambiguity between speaker's reference and semantic reference in various ways. Their findings tentatively confirm that Chinese tend to have genuine descriptivist intuitions about the reference of proper names, while Americans tend to have causal-historical intuitions. Here we report only one of their manipulations.

Suppose that Americans and Chinese alike tend to have causal-historical intuitions about the reference of proper names. Suppose also that the minority of American participants and the majority of Chinese participants in Machery et al. (2004) who answered that John is talking about the man who discovered the incompleteness of arithmetic when he uses "Gödel" did so because they reported an intuition about whom John intends to be talking about. That is, their answer expressed an intuition about speaker's reference, not semantic reference. Then, if Chinese and American participants were given a version of the Gödel case in which John has the intention to refer to the man who stole the theorem, a very large majority of participants should say that John is talking about the man who stole the theorem: Participants whose answer expresses an intuition about semantic reference should answer that John is talking about the theorem stealer since by hypothesis they have causal-historical intuitions, while participants whose answer expresses an intuition about speaker's reference should also answer that John is talking about the theorem stealer because in the vignette John has the intention to talk about the theorem stealer.

To test this prediction, Deutsch and colleagues designed a vignette that began with the original Gödel story, and then continued,

> One night, John is sitting in his room, revising for his mathematics exam by going over the proof of the incompleteness theorem. After a while, he says to his roommate, "Gödel probably got a huge number of awards from mathematical societies!"
>
> When John utters that sentence, he is talking about:
>
> **(A)** the person who got hold of the manuscript and claimed credit for the work.
> **(B)** the person who really discovered the incompleteness of arithmetic.

In this vignette, it is clear that John intends to talk about the man who in fact stole the theorem. Naturally, John does not intend to talk about the man who stole the

theorem under the description "the man who stole the theorem," but it remains true that he intends to talk about this individual since, given that the discoverer of the proof died right after having discovered the proof, only the theorem stealer is likely to have won numerous awards from mathematical societies. So, proponents of the speaker's reference / semantic reference objection should expect a very large majority of participants to answer A. However, Deutsch and colleagues found that, when presented with this vignette, about 75 percent of American participants but only 50 percent of Chinese participants, answered that the speaker is talking about the man who stole the theorem. Because 50 percent of Chinese participants and 25 percent of American participants answered B, one can conclude that it is not the case that everybody has causal-historical intuitions about the semantic reference of proper names and that Americans or Chinese who give answers in line with descriptivist theories of reference are not merely reporting their intuitions about speaker's reference. At the very least, these findings show that genuine intuitions about semantic reference vary among Americans and among Chinese. Furthermore, they undermine the criticisms of the cross-cultural work on intuitions that appeal to the distinction between speaker's and semantic reference to defend the idea that across cultures people tend to have causal-historical intuitions.

We conclude, first, that the findings concerning the intuitions about the reference of proper names cannot be explained away by appealing to the distinction between speaker's reference and semantic reference. Second, and most important for our purposes in this chapter, far from undermining our call for an experimental turn in the philosophy of language, this objection can best be evaluated experimentally: it calls for more and more subtle experiments.

4 Implications of the Experimental Turn

In this final section, we examine the implications the experimental turn might have for the philosophy of language. For the sake of simplicity, we focus on the implications of the findings about variation in intuitions about the reference of proper names, and we argue that philosophers of language face a dilemma: If they acknowledge that ordinary speakers have genuine intuitions about semantic reference that are typically reliable, then in light of the findings discussed in the two previous sections, they need to revise their theories of reference dramatically. However, if they deny that ordinary speakers have such intuitions and reject our call for an experimental turn in the philosophy of language, they then open the door to a challenge about the reality of semantic reference. We further hypothesize that an experimental turn in the philosophy of language will have equally significant implications for other philosophical issues.

Do ordinary speakers of (say) English have genuine intuitions about semantic reference? That is, are they able to identify the semantic reference of proper names in actual and hypothetical situations without confusing it with the speaker's reference of these names in these situations?

Suppose, first, that ordinary speakers of English do have genuine intuitions about semantic reference that are usually reliable. We have reviewed a growing body of experimental evidence that hypothetical situations elicit different answers both within and across cultures, and we have argued that participants' answers in these experiments do reflect their intuitions about semantic reference. This suggests that the reference

of proper names might be determined differently for different competent speakers of a given language (English in the current case) and that descriptivist reference determination might be more common in some cultures than in others. If this is correct, theories of reference should then be modified to account for such variation in reference determination. Reimer (2009) is a pioneering example of a theory of the reference of proper names that is responsive to the variation in intuitions about the reference of proper names—both within and across cultures—illustrated in Sections 2 and 3.

Suppose however that philosophers of language dismiss the arguments presented in Section 3 and maintain that lay people do not have intuitions about semantic reference, perhaps because they do not distinguish between speaker's reference and semantic reference. In this case, ordinary speakers' intuitions can't be used to decide among competing theories of reference proposed by philosophers of language, and surveying them would be a waste of time. But if this is the case, philosophers of language need to clarify the grounds for positing that proper names actually have a semantic reference in addition to the reference that occurrences of names have as a result of speakers' communicative intentions, much as scientists do when they introduce a theoretical distinction between two kinds of properties (for example, *mass* and *weight*). Philosophers of language must tell us what phenomena are best explained by postulating that proper names refer in virtue of the rules of the language they belong to.

We suspect that many philosophers of language would reply that philosophers of language's own intuitions justify positing this distinction: by virtue of their acquired expertise, they might reply, philosophers of language have acquired new intuitions about the reference of proper names, and these intuitions suggest that proper names can have two distinct kinds of reference, speaker's reference and semantic reference (Ludwig 2007; Devitt 2011). Philosophers of language might go on to note that experts (e.g., radiologists) typically acquire new, reliable intuitions by virtue of their acquired expertise (Devitt 2006; Devitt 2011).

We are prepared to concede that many philosophers of language probably do have intuitions consistent with a distinction between speaker's reference and semantic reference. However, for three distinct reasons, we doubt that if ordinary competent speakers do not have genuine intuitions about semantic reference, philosophers' intuitions can justify the hypothesis that proper names have a semantic reference in addition to the reference their occurrences have as a result of speakers' communicative intentions. First, this appeal to the intuitions of experts stands in stark contrast to the growing skepticism in generative linguistics about intuitions that are shared only by linguists. It is curious to see philosophers embrace a methodology (relying on theoreticians' intuitions while discarding lay speakers' intuitions) that linguists are now increasingly questioning. Second, in many fields, experts' intuitions (for example, radiologists' judgments about x-rays) can be validated by objective criteria, and our confidence in the reliability of their intuitions largely depends on such validation (for review, see Weinberg et al. 2010). Since it is (to put it mildly) less than clear what could validate the intuitions of philosophers of language about the reference of proper names, one wonders why these intuitions should be trusted. Third, the broader the consensus among experts, the more credible their intuitions. In this respect, philosophers' intuitions about reference are also on shaky grounds. Some influential linguists and philosophers of language have denied that names *have* a semantic reference, and it is hardly likely that these experts share the intuitions that support postulating that proper names have a semantic reference in addition to speaker's reference (for example, Strawson 1950; Chomsky 1995, 2000; Pietroski 2003). It is

particularly notable that skepticism about referential semantics has long been a key theme in Chomsky's work. In *New Horizons in the Study of Language* he writes (2000: 17):

> In general, a word, even of the simplest kind, does not pick out an entity of the world, or of our "belief space". Conventional assumptions about these matters seem to me very dubious.

Further on in the book Chomsky elaborates on his skepticism:

> A good part of contemporary philosophy of language is concerned with analyzing alleged relations between expressions and things, often exploring intuitions about the technical notions "denote," "refer," "true of," etc. said to hold between expressions and something else. But there can be no intuitions about these notions, just as there can be none about "angular velocity" or "protein." These are technical terms of philosophical discourse with a stipulated sense that has no counterpart in ordinary language. . . . If we rerun the thought experiments with ordinary terms, judgments seem to collapse or, rather, to become so interest-relative as to yield no meaningful results. . . . [I]t is not at all clear that the theory of natural language and its use involves relations of "denotation," "true of," etc. in anything like the sense of the technical theory of meaning.
>
> (Chomsky 2000: 130)

Thus, not only is the reliance of philosophers of language on their own intuitions a departure from the methodological developments in linguistics, there is also little reason to trust philosophers' intuitions since they are not externally validated and they are unlikely to be shared by some of the most prominent linguists.

To recap, if philosophers of language dismiss the type of cross-cultural studies surveyed earlier on the grounds that ordinary speakers don't have genuine intuitions about semantic reference, it is then unclear how they can justify positing the distinction between speaker's reference and semantic reference. But then why should we believe that names in English (and, *mutatis mutandis*, other kinds of words) have a reference in virtue of the rules of English, in addition to the reference their occurrences have as a result of speakers' communicative intentions? On grounds of parsimony, it would seem that one should follow Chomsky's lead and construct our theory of natural language without invoking semantic reference.

Thus, philosophers of language are faced with a dilemma. If they acknowledge that ordinary speakers have intuitions about semantic reference, then theories of reference need to be substantially modified to accommodate the variation in reference determination, as Reimer (2009) has proposed to do. If they deny that ordinary speakers have intuitions about semantic reference, then since philosophers' intuitions do not seem sufficient to justify the assumption that proper names have a semantic reference, it becomes unclear why we should hold that names, instead of people, refer.

Conclusion

It is time for philosophers of language to emulate linguists in taking an experimental turn. The objections against linguists' appeal to their own intuitions carry over to

the methods currently dominant in the philosophy of language, and the burgeoning research in experimental philosophy has made the limits of these methods increasingly clear. Philosophers of language who cling to traditional methods are depriving themselves of a rich source of evidence about how language actually works.

Bibliography

Aoun, J., Hornstein, N., Lightfoot, D. & Weinberg, A. (1987). "Two Types of Locality," *Linguistic Inquiry* 18: 537–77.

Bard, E., Robertson, D. & Sorace, A. (1996). "Magnitude Estimation of linguistic acceptability," *Language* 72: 32–68.

Bauer, L. (1990). "Be-heading the Word," *Journal of Linguistics* 26: 1–31.

Braisby, N., Franks, B. & Hampton, J. (1996). "Essentialism, Word Use, and Concepts," *Cognition* 59: 247–74.

Chomsky, N. (1957). *Syntactic Structures*, The Hague: Mouton.

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*, Chicago: University of Chicago Press.

Chomsky, N. (1976). "Conditions on Rules of Grammar," *Linguistic Analysis* 2: 303–51.

Chomsky, N. (1981). *Lectures on Government and Binding*, Foris: Dordrecht.

Chomsky, N. (1995). "Language and Nature," *Mind* 104: 1–61.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*, Cambridge: Cambridge University Press.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*, Thousand Oaks, CA: SAGE Publications.

Culbertson, J. & Gross, S. (2009). "Are Linguists Better Subjects?", *British Journal for the Philosophy of Science* 60: 721–36.

Deutsch, M. (2009). "Experimental Philosophy and the Theory of Reference," *Mind & Language* 24: 445–66.

Deutsch, M. (2010). "Intuitions, Counter-Examples, and Experimental Philosophy," *The Review of Philosophy and Psychology* 1, 447–460.

Devitt, M. (1981). *Designation*, New York: Columbia University Press.

Devitt, M. (2006). "Intuitions in Linguistics," *The British Journal for the Philosophy of Science* 57: 481–513.

Devitt, M. (2011). "Experimental Semantics," *Philosophy and Phenomenological Research* 82, 418–435.

Evans, G. (1973). "The Causal Theory of Names," *Supplementary Proceedings of the Aristotelian Society* 47: 187–208.

Fanselow, G. (2007). "Carrots—Perfect as Vegetables, but Please Not as a Main Dish," *Theoretical Linguistics* 33: 353–67.

Featherston, S. (2007). "Data in Generative Grammar: The Stick and the Carrot," *Theoretical Linguistics* 33: 269–318.

Genone, J. & Lombrozo, T. (forthcoming). "Concept Possession, Experimental Semantics, and Hybrid Theories of Reference," *Philosophical Psychology*.

Goodall, G. (2006). "Contraction," In M. Everaert and H. van Riemsdijk (eds.) *The Blackwell Companion to Syntax, vol. 1* (pp. 688–703), Malden, MA: Blackwell.

Haider, H. (2007). "As a Matter of Facts—Comments on Featherston's Sticks and Carrots," *Theoretical Linguistics* 33: 381–94.

Hampton, J. A., Estes, Z. & Simmons, S. (2007). "Metamorphosis: Essence, Appearance, and Behavior in the Categorization of Natural Kinds," *Memory & Cognition* 35: 1785–800.

Ichikawa, J., Maitra, I. & Weatherson, B. (Forthcoming). "In Defense of a Kripkean Dogma," *Philosophy and Phenomenological Research*.

Jackendoff, R. S. (1997). *The Architecture of the Language Faculty*, Cambridge, MA: MIT Press.

Jackman, H. (2009). "Semantic Intuitions, Conceptual Analysis, and Cross-Cultural Variation," *Philosophical Studies* 146: 159–77.

Jackson, F. (1998). "Reference and Description Revisited," *Philosophical Perspectives* 12: 201–18.

Jylkkä, J., Railo, H. & Jussi, H. (2009). "Psychological Essentialism and Semantic Externalism: Evidence for Externalism in Lay Speakers' Language Use," *Philosophical Psychology* 22: 37–60.

Keller, F. & Asudeh, A. (2001). "Constraints on Linguistic Coreference: Structural vs. Pragmatic Factors," In J. D. Moore & K. Stenning (eds.) *Proceedings of the 23rd. Annual Conference of the Cognitive Science Society* (pp. 483–88), Mahwah, NJ: Lawrence Erlbaum Associates.

Kripke, S. (1972/1980). *Naming and Necessity*, Cambridge, MA: Harvard University Press.
Kripke, S. (1977). "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy* 2: 255–76.
Labov, W. (1975). "Empirical Foundations of Linguistic Theory," In R. Austerlitz (ed.) *The Scope of American Linguistics* (pp. 77–133), Lisse: Peter de Ridder.
Labov, W. (1996). "When Intuitions Fail," *Chicago Linguistics Society* 32: 77–105.
Lam, B. (2010). "Are Cantonese Speakers Really Descriptivists? Revisiting Cross-Cultural Semantics," *Cognition* 115: 320–9.
Lasnik, H. & Saito, M. (1984). "On the Nature of Proper Government," *Linguistic Inquiry* 15: 235–89.
Ludwig, K. (2007). "The Epistemology of Thought Experiments: First-Person Approach vs. Third-Person Approach," *Midwest Studies in Philosophy* 31: 128–59.
Machery, E. (2011). "Variation in Intuitions about Reference and Ontological Disagreements," In S. D. Hales (ed.) *A Companion to Relativism*, (pp. 118–36) New York: Blackwell.
Machery, E., Deutsch, M. & Sytsma J. (forthcoming). "Speaker's Reference and Cross-Cultural Semantics." In A. Bianchi (ed.), *On Reference*, Oxford University Press.
Machery, E., Mallon, R., Nichols, S. & Stich, S. P. (2004). "Semantics, Cross-Cultural Style," *Cognition* 92: B1–12.
Machery, E., Mallon, R., Nichols, S., Stich, S. P., Deutsch, M. & Sytsma, J. (2010) "Semantic Intuitions: Reply to Lam," *Cognition* 115: 320–29.
Machery, E., Olivola, C. Y. & de Blanc, M. (2009). "Linguistic and Metalinguistic Intuitions in the Philosophy of Language," *Analysis* 69: 689–94.
Marantz, A. (2005). "Generative Linguistics within the Cognitive Neuroscience of Language," *The Linguistic Review* 22, 429–445
Martí, G. (2009). "Against Semantic Multi-Culturalism," *Analysis* 69: 42–8.
Myers, J. (2009). "Syntactic Judgment Experiments," *Language and Linguistics Compass* 3: 406–23.
Nisbett, R. E. (2003). *The Geography of Thought: how Asians and Westerners Think Differently . . . And Why*, New York: Free Press.
Novack, I. A. & Reboul, A. (2008). "Experimental Pragmatics: A Gricean Turn in the Study of Language," *Trends in Cognitive Sciences* 12: 425–31.
Nunberg, G., Sag, I. A. & Wasow, T. (1994). "Idioms," *Language* 703: 491–538.
Phillips, C. (2009). "Should we Impeach Armchair Linguists?," In S. Iwasaki (ed.), *Japanese/Korean Linguistics* 17, Palo Alto, CA: CSLI Publications.
Pietroski, P. (2003). "The Character of Natural Language Semantics," In A. Barber (ed.) *Epistemology of Language* (pp. 217–56), New York: Oxford University Press.
Pietroski, P., Lidz, J., Hunter, T. & Halberda, J. (2009). "The Meaning of 'Most': Semantics, Numerosity, and Psychology," *Mind & Language* 24: 554–85.
Putnam, H. (1970). "Is Semantics Possible?" In H. Kiefer and M. Munitz (eds.) *Languages, Belief and Metaphysics* (pp. 139–52), Albany: SUNY Press.
Putnam, H. (1975). "The Meaning of 'Meaning'," In H. Putnam (ed.), *Philosophical Papers*, *Vol. 2: Mind, Language, and Reality* (pp. 215–71), New York: Cambridge University Press.
Reimer, M. (2009). "Jonah Cases," In A. Everett (ed.) *Empty Names*, Oxford: Oxford University Press.
Riemer, N. (2009). "Grammaticality as Evidence and as Prediction in a Galilean Linguistics," *Language Sciences* 31: 612–33.
Riehemann, S. Z. (2001). A Constructional Approach to Idioms and Word Formation, Stanford University Dissertation.
Sampson, G. (2005). "Quantifying the Shift Towards Empirical Methods," *International Journal of Corpus Linguistics* 10: 15–36.
Schütze, C. T. (1996). *The Empirical Base of Linguistics*, Chicago: The University of Chicago Press.
Schütze, C. T. (2005). "Thinking About What We Are Asking Speakers to Do," In S. Kepser and M. Reis (eds.) *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives* (pp.457–485), Berlin: Mouton de Gruyter.
Searle, J. (1958). "Proper Names," *Mind* 67: 166–73.
Sorace, A. & Keller, F. (2005). "Gradience in Linguistic Data," *Lingua* 115: 1497–524.
Sprouse, J. (2007). A Program for Experimental Syntax: Finding the Relationship Between Acceptability and Grammatical Knowledge, College Park, MD: University of Maryland dissertation.
Stich, S. P. (2010). "Philosophy and WEIRD Intuition," *Behavioral and Brain Sciences* 33:110–11.
Strawson, P. F. (1950). "On Referring," *Mind* 59: 320–44.

Sytsma, J. M. & Livengood, J. (2011). "A New Perspective Concerning Experiments on Semantic Intuitions," *The Australasian Journal of Philosophy* 89: 315–32.
Wasow, T. & Arnold, J. (2005). "Intuitions in Linguistic Argumentation," *Lingua* 115: 1481–96.
Weinberg, J. M., Gonnerman, C., Buckner, C. & Alexander, J. (2010). "Are Philosophers Expert Intuiters?", *Philosophical Psychology* 23: 331–55.
Wikforss, Å. (2008). "Semantic Externalism and Psychological Externalism," *Philosophy Compass* 3: 158–81.

4.2
THE ROLE OF LINGUISTICS

Sarah Moss

The interaction of linguistics and the philosophy of language regularly inspires florid description. One philosopher describes his discipline as the "initial central sun, seminal and tumultuous: from time to time it throws off some portion of itself to take station as a planet," and he goes on to suggest that linguistics may be among its satellites (Austin 1956, 232). In the words of one linguist, philosophy is more like "a rummage sale that sometimes offers useful items for the home improver" (Nunberg 2002, 679). Higginbotham 2002 offers the tantalizingly ambiguous observation that "linguistics and philosophy, like steak and barbecue sauce, have much to give each other" (575). And contemporary interactions between linguists and philosophers of language are so extensive that these descriptions fail to capture the extent to which it is increasingly difficult just to delineate the fields.

Following the maxim that it would be "better to show than tell" something about the role of linguistics in the philosophy of language, I will focus on a few case studies that illustrate key aspects of this relationship. This entry is necessarily limited in scope. Partee 2004b and 2005 provide an excellent account of early interactions between linguists and philosophers, so this entry focuses on contemporary literature rather than the historical role of linguistics in the philosophy of language. Furthermore, I set aside the extensive influence of linguistics on other subfields of philosophy, such as the influence of psycholinguistics in the philosophy of mind, the semantics of knowledge constructions in assessing *intellectualism* in epistemology (cf. Stanley & Williamson 2001), and the semantics of event-related counting in the metaphysics of persistence (cf. Moss 2010b). I focus on the relationship between philosophy of language and three branches of linguistics—syntax, semantics, and pragmatics. In each section, I discuss a case study in detail and then briefly mention other notable instances of the role of linguistics in the philosophy of language.

1 Syntax

The first case study is addressed most directly in Stanley 2000: the debate over whether there are in fact *unarticulated constituents* of the propositions expressed by utterances. For instance, it is uncontroversial that both (1) and (2) are context-sensitive, and that the place where they are uttered is relevant to their truth conditions:

(1) They are serving drinks at the bar near here.
(2) They are serving drinks at the local bar.

Some have argued that unlike (1), sentences such as (2) do not contain any element that designates the place of the context of utterance (cf. Pragmatic Enrichment for references and an assessment of this debate). Instead there is a "post-semantic" level of interpretation in which such unarticulated constituents of the proposition are identified. Perry 1986 introduces the notion of an unarticulated constituent, and Perry 1998 clarifies:

> In some cases of implicit reference there is a feature, a trace, a sort of phantom expression, that serves in place of an expression, so the referred to constituent really isn't unarticulated . . . I am interested in the theoretical possibility and coherence of truly unarticulated constituents. (9)

Against Perry, Stanley 2000 argues that sentences such as (3) are evidence against the unarticulated constituent analysis:

(3) Every newspaper reporter went to a local bar to hear the news.

Stanley argues that (3) contains a covert location variable that is semantically and hence syntactically bound on the natural reading of (3), and that is *a fortiori* present as a syntactically articulated element of the sentence.

Stanley also suggests that covert pronouns associated with expressions such as 'local' are responsible for *weak crossover effects*, restrictions on binding easily observed in sentences containing overt pronouns. For instance, one cannot hear a reading of (4) on which the overt pronoun 'her' is bound by the lower quantifier 'every reporter':

(4) #Her$_i$ local bar sponsored [every reporter]$_i$.

Stanley argues that one can see the same effect in (5), namely the absence of a reading that says that every reporter was sponsored by a bar local to that reporter:

(5) #A local bar sponsored every reporter.

In other words, the same syntactic constraints on binding seem responsible for the absence of the intended bound readings of (4) and (5) (though see Rothschild & Segal 2009 for dissent regarding such examples). Stanley concludes that the behavior of relational expressions (e.g. 'local') in weak crossover constructions such as (5) "strongly suggests the existence of a covert pronominal element in relational expressions" (423).

This is a classic case of syntactic arguments playing a role in the philosophy of language. The arguments in fact originate with linguists: Partee 1989 cites Mitchell 1986 as the first to identify a "much broader class of contentful context-dependent elements which can also exhibit bound-variable-like behavior, such as *local*" (260). Partee also

points out weak crossover effects in sentences like (6), on the reading where each professor's local union is in question:

(6) #?The leader of the local union wrote a letter to every untenured professor.

Condoravdi & Gawron 1996 give similar examples and argue that such examples demonstrate that "the contextual determination of the interpretation of implicit arguments is not just a matter of pragmatic principles" (6). These linguists explicitly mention Crimmins 1992 as an advocate of unarticulated constituents, and as such they take him to be one philosopher targeted by their analysis.

It is perhaps regrettable that philosophers engaged in this dispute seldom mention that Partee 1989 and Condoravdi & Gawron 1996 ultimately reject the pronominal analysis of relational expressions defended in Stanley 2000. In particular, Condoravdi & Gawron 1996 argue that "implicit arguments pattern with definite descriptions rather than pronouns in allowing for associative anaphoric readings" (10). Partee 2009 notes that semanticists rarely advocate the covert pronoun analysis argued against in Partee 1989. But she mentions that semanticists do accept the main point of Stanley 2000:

> Formal semanticists are quite uniformly convinced that the Pragmatic Enrichment Approach cannot be right, because it does not predict any purely linguistic constraints on possible enrichment, and there are indeed strong anaphora-like constraints on what is possible. (8)

It is important for the opponent of pragmatic enrichment to consider a range of responses to the above syntactic arguments. For instance, one might claim that the effects of pragmatic enrichment may be semantically controlled by other expressions. See Cappelen & Lepore 2002, 2005 for responses to the above arguments and Cohen & Rickless 2007 for a rejoinder, and see Marti 2006 and Elbourne 2008 for further references and critical discussion of both camps. Regardless of one's opinion about unarticulated constituents, it is clear that syntactic arguments have played a major role in this dispute.

Furthermore, the arguments just considered have consequences for other debates in the philosophy of language, far removed from discussions of pragmatic enrichment. For instance, take a topic that has been part of the philosophy of language canon since Frege 1892: the correct theory of attitude ascriptions. Concerns raised by direct reference theorists (e.g. Kripke 1972, Putnam 1975) motivate many philosophers to claim that coreferential names have the same semantic value. Recent advocates of this doctrine give various proposals for reconciling two contrasting claims: first, that coreferential names have the same semantic value even when embedded in intensional contexts, and second, that sentences such as (7) and (8) can nevertheless have different truth conditions:

(7) Lois believes that Superman is strong.
(8) Lois believes that Clark Kent is strong.

Furthermore, (9) can have different truth conditions in different contexts, depending on which way of thinking of Paderewski is salient (cf. Kripke 1979):

(9) Peter believes that Paderewski had musical talent.

Crimmins 1992 argues that different guises of Paderewski are unarticulated constituents of propositions expressed by (9) in different contexts, while Schiffer 1992 discusses the proposal that modes of presentation are covert implicit arguments of attitude verbs (cf. Ludlow 1996 and Schiffer 1996 for syntactically driven assessments of hidden-indexical theories of attitude ascriptions). Richard 1990 takes a third approach: that 'believes' in (9) denotes different belief relations in contexts in which different sentences count as acceptable translations of sentences Peter accepts.

The same arguments marshalled against the theory that locations are unarticulated constituents of propositions expressed by 'local' sentences can be brought to bear on the theory that guises are unarticulated constituents of propositions expressed by 'believes' sentences. Consider the following case:

> You are approached by Cicero, a man you recognize to be a famous spy. Cicero says, "I see you know me as a notorious spy. But you also know me as one of your closest friends, and without realizing that I am a spy, you share secrets with me as with any trusted person. And the same goes for every woman in this room: each knows me as one of her closest friends, and none yet realizes that I am a spy."

In such a case, you may truly say:

(10) No woman in this room yet realizes that Cicero is a spy.

Arguably this utterance has the following reading:

(11) No woman in this room yet realizes that Cicero is a spy when thinking about Cicero in the one-of-her-closest-friends way.

But this bound reading tells against the unarticulated constituent analysis of (10) and other attitude ascriptions as much as the bound reading of (3) tells against the unarticulated constituent analysis of 'local' sentences. Stanley 2000 generalizes his argument to refute a narrow indexical analysis of relational expressions. I leave it to the reader to decide if the analogous variant of the above argument presents an equal challenge for the narrow indexical semantics for attitude ascriptions defended in Richard 1990. Here the central point is simply that syntactic arguments are relevant in the context of an entirely self-sufficient philosophical debate.

There are a range of other notable instances of syntactic arguments bearing on classic questions in the philosophy of language. Several arguments regarding quantifier scope have been applied in assessing quantificational analyses of definite descriptions; see Glanzberg 2007 for critical discussion of some of this literature. Others have applied syntactic arguments in assessing quantificational accounts of complex demonstratives; see King 2001 for one example and Lepore & Johnson 2002 and King 2010 for critical discussion. And it has been suggested that the appropriate syntactic representation of tense has implications for the nature of the objects of assertion, e.g. for whether such objects vary in truth value across times; see Ninan 2010 for a critical discussion of the role that syntactic arguments play in this literature.

2 Semantics

A second case study for the influence of linguistics in the philosophy of language is presented in the discussion of subjunctive conditionals in von Fintel 2001. Giving an analysis of subjunctives has been of philosophical interest since Chisholm 1946 and Goodman 1947. Stalnaker 1968 and Lewis 1973 advocate competing theories of subjunctives, and many others pursue details of their dispute (see Subjunctive Conditionals). But the competing accounts given by these philosophers do share a central feature: according to both, Antecedent Strengthening is invalid. For instance, consider the following felicitous sequence from Lewis 1973:

(12) a. If the USA threw its weapons into the sea tomorrow, there would be war.
b. If the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace.

Stalnaker and Lewis both advocate *variably strict* conditional analyses, according to which 'if *p*, would *q*' expresses a proposition that is true roughly just in case all the closest *p* worlds are *q* worlds. In particular, Lewis 1973 says that 'if *p*, would *q*' expresses a proposition that is non-vacuously true at a world just in case some *p*-and-*q* world is closer to that world than any *p*-and-not-*q* world. Stalnaker 1968 says that 'if *p*, would *q*' expresses a proposition that is true at a world just in case the closest *p* world is a *q* world. Both predict that strengthening the antecedent of (12-a) need not preserve truth, so (12-a) and (12-b) can both be true.

Recent papers in formal semantics challenge the variably strict conditional analysis. In von Fintel 1999, 2001, von Fintel argues that Stalnaker and Lewis fail to explain the infelicity of other sequences such as:

(13) a. If the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace.
b. #If the USA threw its weapons into the sea tomorrow, there would be war.

In order to account for the asymmetry between (12) and (13), von Fintel endorses a strict conditional analysis (cf. Gillies 2007 for a similar challenge and response). He argues that 'if *p*, would *q*' is true just in case all the *p* worlds in a contextually determined domain are *q* worlds, and he adds that 'if *p*, would *q*' itself demands that the relevant domain contain some *p* worlds. So after a speaker utters (13-a), the context shifts so that the domain of the strict conditional expressed in (13-b) contains worlds where all nuclear powers disarm, and (13-b) therefore fails to be true. In brief, von Fintel argues that he can beat Stalnaker and Lewis at their own game, since the variably strict analysis that does so well at accounting for (12) fails to account for our judgment about (13) (cf. Moss 2010a and Moss 2010c for responses on behalf of the variably strict conditional analysis).

Furthermore, von Fintel is not only interested in accounting for judgments about more creative sequences of subjunctives. He also raises distinctively semantic concerns that bear on the dispute. The most interesting linguistic motivation for his strict conditional analysis comes from facts about negative polarity items (NPIs), expressions that

are licensed only in the scope of negation and other "negative" expressions. English 'ever' is one example:

(14) I haven't ever watched that movie.
(15) #I have ever watched that movie.

Ladusaw 1979 argues that NPIs are licensed in *downward entailing* environments, i.e. environments that reverse entailment relations (such as the relation between 'fat cat' and 'cat'). For example, 'every' is downward entailing in just its first argument and 'no' is downward entailing in both, as evidenced by the following entailment relations:

(16) Every cat sat. $\Rightarrow$ Every fat cat sat.
(17) Every cat sat. $\not\Rightarrow$ Every cat sat on the mat.
(18) No cat sat. $\Rightarrow$ No fat cat sat.
(19) No cat sat. $\Rightarrow$ No cat sat on the mat.

And 'ever' is licensed in the environments that are downward entailing (cf. von Fintel 1999 for a catalog of literature disputing and refining this generalization):

(20) Every boy who ever watched that movie liked it.
(21) #Every boy who watched that movie ever liked it.
(22) No boy who ever watched that movie liked it.
(23) No boy who watched that movie ever liked it.

Furthermore, antecedents of conditionals—including subjunctives—license NPIs:

(24) If you had ever watched that movie, you would have liked it.

In 2001, von Fintel argues that this fact is evidence for the strict conditional analysis of subjunctives. Recall that on the variably strict conditional account, antecedents of subjunctives are not downward entailing environments, i.e. Antecedent Strengthening is semantically invalid. But the strict conditional acccount holds that antecedents are downward entailing, as the behavior of (24) suggests. One might respond that superlatives also license NPIs, as in 'the oldest boy who ever watched that movie liked it'. Fans of the variably strict analysis could argue that a superlative element (e.g. 'closest') in the logical form of subjunctives licenses NPIs in sentences like (24). See von Fintel 1999 for a critique of this proposal. Whatever one may make of this debate, it is clear that facts about negative polarity items present further data for any philosopher developing a theory of subjunctive conditionals.

Besides tests involving negative polarity items, several tools from semantics regularly appear in the philosophy of language literature. Philosophers use minimal pairs of sentences in constructing and testing hypotheses. Discussions of verb phrase ellipsis and constraints on binding are scattered throughout the literature. Cross-linguistic studies have introduced constraints on philosophical theories of language fragments (e.g. connections between research on evidentials and epistemic modals, and discussions of non-English demonstratives outlined in Wolter 2009). Tests for distinguishing predicational from specificational copular sentences have played a role in debates about the referents of 'that'-clauses (cf. Pryor 2007). These examples suggest that perhaps semantics also is a source of useful items for the home improver.

In addition to providing philosophers with useful tools, semantics is a source of research interests common to linguists and philosophers of language. Semanticists have studied literally every topic in the first section of this volume (cf. Parts of Speech). Obviously there has been much shared research on the semantics of various referring expressions, including anaphoric expressions (cf. Anaphora), indefinite and definite descriptions (cf. Descriptions), and complex demonstratives (cf. Indexicals and Demonstratives). Semantic analyses of other language fragments have implications for foundational questions about the nature of assertion, e.g. research on epistemic modals (cf. Swanson 2008 for references) and other language of subjective uncertainty and predicates of personal taste (cf. Relativism). Philosophers have also actively engaged in linguistics research at the developmental stage of novel semantic frameworks, e.g. from the development of formal semantic systems (Lewis 1970, Montague 1970, 1973) to event semantics (Davidson 1967), double-indexing (Kamp 1971, Kaplan 1977), alternative semantics (Hamblin 1973), dynamic semantics (Stalnaker 1974, 1978, Lewis 1975, 1979a, 1979b, Kamp 1981, Veltman 1996), situation semantics (Barwise & Cooper 1981), and structured propositions (Cresswell & Stechow 1982). In each instance, philosophers have contributed to the development of research areas squarely situated in the field of formal semantics.

3 Pragmatics

In a number of cases, philosophers have also been present in the early stages of research programs in pragmatics. Recently linguists (e.g. Potts 2005) have developed more formal approaches to theories of speech acts first introduced by Grice 1987, and the same can be said for early discussions of presupposition in Stalnaker 1970, 1974, Lewis 1979b and Kripke 2009 (originally delivered as a lecture in 1990). The third case study I shall discuss involves applying some more recent literature on presupposition in giving an account of the following classic pair of sentences:

(25) Hesperus is Hesperus.
(26) Hesperus is Phosphorus.

Since the introduction of these sentences in Frege 1892, philosophers of language have confronted the challenge of accounting for the fact that sentences differing only by the substitution of coreferential names can nevertheless convey different information.

Several direct reference theorists have argued that pragmatic differences between (25) and (26) are responsible for their difference in informativeness (cf. Salmon 1986, Soames 2002, Thau 2002). Recent considerations raised by linguists confirm this general strategy but suggest a novel pragmatic vehicle for the content conveyed in (26), namely *presupposed* contents. Here again the account is informed by empirical considerations, namely a relatively straightforward test battery for presuppositions. For example, possessives typically introduce presuppositions that project through negation and antecedents of conditionals, so that each of the following sentences conveys that John has a dog (see Ch. 1 of Geurts 1999 for more detailed discussion of projection tests for presuppositions):

(27) John's dog is hungry.
(28) John's dog is not hungry.
(29) If John's dog is hungry, then John should feed him.

But presuppositions can also be locally accommodated in the consequent of a conditional. Such conditionals will therefore fail to convey the same presupposition, e.g. the following does not convey that John has a dog:

(30) If John has a dog, John's dog is hungry.

Several theorists have argued that names convey contents that exhibit similar behavior. For instance, the following are only appropriate if the name 'Fido' meets some familiarity condition (cf. Sommers 1982 for an early instance of this observation):

(31) Fido is hungry.
(32) Fido is not hungry.
(33) If Fido is hungry, then someone should feed him.

But this condition of familiarity can be met by a local context, so that e.g. the speaker of (34) does not presuppose that the name 'Fido' is familiar to his audience:

(34) If Fido is that pink poodle over there, then Fido is hungry.

Maier 2009 argues that "dynamic semantics and presupposition theory have given rise to a fruitful analysis of proper names as presupposition inducers" (265).

Details of presuppositional theories of names vary widely. Geurts 1997 argues that "the meaning of a name *N* is 'the individual named *N*', where the semantic contribution of the definite article is to be analysed in presuppositional terms" (340). Swanson 2006 claims that the speaker of (31) presupposes merely that "the thing she associates with 'Fido' in the context of utterance is the thing the addressee associates with 'Fido' in that context" (11). Maier 2009 operates in a framework according to which names introduce discourse referents that are themselves objects of presupposition. Roberts 2009 argues that names are special because the discourse referent that satisfies the familiarity presupposition of a name "carries the information that the individual bearer. . . does so by virtue of a socially recognized, causal dubbing event by an (authorized) agent, and the subsequent social propagation of the name-individual association. . . to the current occasion of use" (7). But these disparate accounts are unified in spirit: they say that sentences such as 'Hesperus is Phosphorus' convey presupposed information that distinguishes them from trivial identity sentences.

Direct reference theorists are interested in the difference in the cognitive import of (25) and (26). In explaining this difference, they also aim to give a theory of our judgments about attitude ascriptions:

(35) John believes that Hesperus is Hesperus.
(36) John believes that Hesperus is Phosphorus.

So we have arrived where we began: Insofar as the original syntactic arguments outlined in §1 are dispositive, general pragmatic theories of differences in cognitive import are constrained by syntactic considerations. For instance, they must accommodate bound readings of ascriptions such as (10):

(10) No woman in this room yet realizes that Cicero is a spy.

Providing for these considerations is compatible with a pragmatic account. Presuppositions introduced by names may be locally accommodated in attitude ascriptions. Stalnaker 1988 introduces the notion of a "derived context" corresponding to embedded clauses in ascriptions, and there are other formal variations of this notion (cf. Swanson 2010 for an alternative approach and further comparisons between names and other presupposition-inducing expressions). In the course of giving any such account, it will be an important fact that locally accommodated presuppositions can be bound by higher operators. For instance, roughly speaking, (37) presupposes that every woman in the room believes that she herself already met John:

(37) No woman in this room believes that she will meet John again.

For pragmatic theorists, similar readings of locally accommodated presuppositions are apt to figure in explanations of syntactic considerations raised by bound readings of attitude ascriptions such as (10). Just to take one example, advocates of Discourse Representation Theory may claim it as an advantage that their frameworks easily model such behavior (e.g. see Asher 1986 for a DR-theoretic analysis of puzzles about attitude ascriptions, such as those in Kripke 1979). In any case, one can see that constraints delivered by linguistics may be useful guides to further progress here.

4 Reflections

Semanticists are separated from philosophers of language in virtue of being alert to quite different adjacent fields: morphology, phonology, and syntax as opposed to metaphysics, epistemology, and the philosophy of mind, for instance. As a result, even linguists and philosophers engaging in similar research programs sometimes find different questions interesting and different arguments most useful for theory building. Examples given here illustrate that the different perspective of the linguist is a valuable asset; it is a reason why linguistics offers the philosophy of language so many useful tools for constraining theory development. In this respect, linguistics is like other disciplines discussed in this section of the volume, while unique in the extent to which some of its practitioners and some philosophers are engaged in a common enterprise.

Ultimately the case studies canvassed here point to a simple conclusion: linguistics plays a variety of distinctive roles in the philosophy of language. Linguists have contributed empirical constraints on theories about classic topics in the philosophy of language, in some cases thereby reinvigorating or redirecting otherwise settled debates. Linguists have been in direct dialogue with philosophers as collaborators and opponents of particular theories. And linguists sometimes continue philosophical debates until they may fairly claim to have appropriated that debate as part of linguistics proper. On some occasions, linguists and philosophers of language have operated in parallel on problems, and the role of linguistics in philosophical discussions has not always been as great as it could have been.

Thomason & Pelletier 2002 point out one unique facet of the role of linguistics in philosophy: "the results of linguistics—and especially the results of the branches of linguistics concerned with semantics—provide direct intuitions of the sort that have always served as the starting points of philosophical positions" (509). In light of this observation, it is understandable that the role of linguistics in the philosophy of language is not only varied but malleable. The question of what role linguistics should play is in part a philosophical question about the role of intuitions in philosophical

inquiry, and answers to this question vary according to the philosopher, the particular inquiry under discussion, and the times. The roles played by intuitions in each field are themselves evolving in the twenty-first century as online engines become more refined guides for linguists' corpus searches and as experimental philosophy research expands, inciting metaphilosophical debate. But even as the role of linguistics in the philosophy of language is constantly changing, one may recognize it as a constant fact that the relationship between the disciplines is greatly productive for both.

References

Asher, Nicholas. 1986. "Belief in Discourse Representation Theory." *Journal of Philosophical Logic*, vol. 15 (2): 127–189.

Austin, J. L. 1956. "Ifs and Cans." In *Philosophical Papers*, J. O. Urmson & G. J. Warnock, editors. Oxford University Press, Oxford, first edn.

Barwise, John & Robin Cooper. 1981. "Generalized Quantifiers and Natural Language." In Portner & Partee (2002), 75–126.

Cappelen, Herman & Ernest Lepore. 2002. "Indexicality, Binding, Anaphora, and A Priori Truth." *Analysis*, vol. 62: 271–281.

——. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Blackwell Publishing Ltd., Malden, MA.

Chisholm, Roderick M. 1946. "The Contrary-to-Fact Conditional." *Mind*, vol. 55 (220): 289–307.

Cohen, Jonathan & Samuel C. Rickless. 2007. "Binding Arguments and Hidden Variables." *Analysis*, vol. 67 (1): 65–71.

Condoravdi, Cleo & Jean Gawron. 1996. "The Context-Dependency of Implicit Arguments." In *Quantifiers, Deduction, and Context*, Makoto Kanazawa, Christopher Piñón & Henriëtte de Swart, editors, 1–32. CSLI Publications, Stanford.

Cresswell, Maxwell & Arnim Stechow. 1982. "*De Re* Belief Generalized." *Linguistics and Philosophy*, vol. 5 (4): 503–535.

Crimmins, Mark. 1992. *Talk About Beliefs*. MIT Press, Cambridge.

Davidson, Donald. 1967. "The Logical Form of Action Sentences." In *The Logic of Decision and Action*, Nicholas Rescher, editor, 81–95. University of Pittsburgh Press, Pittsburgh.

Elbourne, Paul. 2008. "The Argument From Binding." *Philosophical Perspectives*, vol. 22 (1): 89–110.

von Fintel, Kai. 1999. "NPI Licensing, Strawson Entailment, and Context Dependency." *Journal of Semantics*, vol. 16: 97–148.

——. 2001. "Counterfactuals in a Dynamic Context." In *Ken Hale: A Life in Language*, Michael Kenstowicz, editor. MIT Press, Cambridge.

Frege, Gottlob. 1892. "On *Sinn and Bedeutung*." In *The Frege Reader*, Michael Beaney, editor, 151–171. Blackwell Publishers, Ltd., Oxford. Translated by Max Black.

Geurts, Bart. 1997. "Good News About the Description Theory of Names." *Journal of Semantics*, vol. 14: 319–348.

——. 1999. *Presuppositions and Pronouns*. Elsevier, Amsterdam.

Gillies, Thony. 2007. "Counterfactual Scorekeeping." *Linguistics and Philosophy*, vol. 30: 329–360.

Glanzberg, Michael. 2007. "Definite Descriptions and Quantifier Scope: Some Mates Cases Reconsidered." *European Journal of Analytic Philosophy*, vol. 3: 133–158.

Goodman, Nelson. 1947. "The Problem of Counterfactual Conditionals." *Journal of Philosophy*, vol. 44 (5): 113–128.

Grice, Paul. 1987. "Logic and Conversation." In *Studies in the Way of Words*. Harvard University Press, Cambridge.

Hamblin, Charles. 1973. "Questions in Montague English." *Foundations of Language*, vol. 10: 41–53.

Higginbotham, James. 2002. "On Linguistics in Philosophy, and Philosophy in Linguistics." *Linguistics and Philosophy*, vol. 25 (5–6): 573–584.

Kamp, Hans. 1971. "Formal Properties of 'Now'." *Theoria*, vol. 37: 227–274.

——. 1981. "A Theory of Truth and Semantic Representation." In *Formal Methods in the Study of Language: Mathematical Centre Tracts* 135, J. Groenendijk, T. Janssen & M. Stokhof, editors, 277–322. Stichting Mathematisch Centrum, Amsterdam.

Kaplan, David. 1977. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In *Themes from Kaplan*, Joseph Almog, John Perry & Howard Wettstein, editors, 481–563. Oxford University Press, Oxford.
King, Jeffrey C. 2001. *Complex Demonstratives: A Quantificational Account*. MIT Press, Cambridge.
——. 2010. "Syntactic Evidence for Semantic Claims." Manuscript, Department of Philosophy, Rutgers University.
Kripke, Saul. 1972. "Naming and Necessity." In *Semantics of Natural Language*, Donald Davidson & Gilbert Harman, editors, 253–355. Reidel, Dordrecht.
——. 1979. "A Puzzle About Belief." In *Readings in the Philosophy of Language*, Peter Ludlow, editor, 875–920. MIT Press, Cambridge.
——. 2009. "Presupposition and Anaphora: Remarks on the Formulation of the Projection Problem." *Linguistic Inquiry*, vol. 40 (3): 367–386.
Ladusaw, William. 1979. Polarity Sensitivity as Inherent Scope Relations. Ph.D. thesis, University of Texas at Austin.
Lepore, Ernest & Kent Johnson. 2002. "Does Syntax Reveal Semantics?: A Case Study of Complex Demonstratives." *Philosophical Perspectives*, vol. 16: 17–41.
Lewis, David K. 1970. "General Semantics." In Lewis (1983), 189–232.
——. 1973. *Counterfactuals*. Basil Blackwell Ltd., Malden, MA.
——. 1975. "Adverbs of Quantification." In *Formal Semantics of Natural Language*, Edward L. Keenan, editor, 3–15. Cambridge University Press, Cambridge U.K.
——. 1979a. "A Problem About Permission." In *Papers in Ethics and Social Philosophy*. Cambridge University Press, Cambridge U.K.
——. 1979b. "Scorekeeping in a Language Game." In Lewis (1983), 233–249.
——. 1983. *Philosophical Papers*, vol. 1. Oxford University Press, Oxford.
Ludlow, Peter. 1996. "The Adicity of 'Believes' and the Hidden Indexical Theory." *Analysis*, vol. 56 (2): 97–101.
Maier, Emar. 2009. "Proper Names and Indexicals Trigger Rigid Presuppositions." *Journal of Semantics*, vol. 26 (3): 253–315.
Marti, Luisa. 2006. "Unarticulated Constituents Revisited." *Linguistics and Philosophy*, vol. 29: 135–166.
Mitchell, Jonathan. 1986. The Formal Semantics Point of View. PhD. dissertation, Department of Linguistics, University of Massachusetts, Amherst.
Montague, Richard. 1970. "Universal Grammar." *Theoria*, vol. 36: 373–398.
——. 1973. "The Proper Treatment of Quantification in Ordinary English." In Portner & Partee (2002), 17–34.
Moss, Sarah. 2010a. "Subjunctive Credences and Semantic Humility" Forthcoming in *Philosophy and Phenomenological Research*.
——. 2010b. "Four-Dimensionalist Theories of Persistence." Forthcoming in *Australasian Journal of Philosophy*.
——. 2010c. "On the Pragmatics of Counterfactuals." Forthcoming in *Noûs*.
Ninan, Dilip. 2010. "Semantics and the Objects of Assertion." *Linguistics and Philosophy*, vol. 33 (5): 335–380.
Nunberg, Geoffrey. 2002. "Do You Know What It Means to Miss New Orleans?" *Linguistics and Philosophy*, vol. 25 (5–6): 671–680.
Partee, Barbara H. 1989. "Binding Implicit Variables in Quantified Contexts." In Partee (2004a), 259–281.
——. 2004a. *Compositionality in Formal Semantics*. Blackwell Publishers, Ltd., Oxford.
——. 2004b. "Reflections of a Formal Semanticist." In Partee (2004a), 1–25.
——. 2005. "Reflections of a Formal Semanticist as of Feb 2005." Manuscript, University of Massachusetts, Amherst.
——. 2009. "Lecture 12: Implicit Arguments and Point of View." Manuscript, University of Massachusetts, Amherst.
Perry, John. 1986. "Thought Without Representation." *Supplementary Proceedings of the Aristotelian Society*, vol. 60: 137–152.
——. 1998. "Indexicals, Contexts, and Unarticulated Constituents." In *Computing Natural Language*, A. Aliseda, R. van Gabeek & D. Westerståhl, editors, 1–11. CSLI Publications, Stanford.
Portner, Paul & Barbara H. Partee, editors. 2002. *Formal Semantics: The Essential Readings*. Blackwell, Oxford.
Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford University Press, Oxford.

Pryor, Jim. 2007. "Reasons and That-Clauses." *Philosophical Issues*, vol. 17 (1): 217–244.
Putnam, Hilary. 1975. "The Meaning of 'Meaning'." In *The Twin Earth Chronicles*, Andrew Pessin & Sanford Goldberg, editors. M. E. Sharpe, Armonk, NY.
Richard, Mark. 1990. *Propositional Attitudes: An Essay on Thoughts and HowWe Ascribe Them*. Cambridge University Press, Cambridge.
Roberts, Craige. 2009. "Names as Definites." Ms., Department of Linguistics, Ohio State University.
Rothschild, Daniel & Gabriel Segal. 2009. "Indexical Predicates." *Mind and Language*, vol. 24 (4): 467–493.
Salmon, Nathan. 1986. *Frege's Puzzle*. MIT Press, Cambridge.
Schiffer, Stephen R. 1992. "Belief Ascription." *Journal of Philosophy*, vol. 89 (10): 499–521.
——. 1996. "The Hidden-Indexical Theory's Logical Form Problem: A Rejoinder." *Analysis*, vol. 56 (2): 92–97.
Soames, Scott. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press, Oxford.
Sommers, F. 1982. *The Logic of Natural Language*. Clarendon, Oxford.
Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Ifs: Conditionals, Belief, Decision, Chance, and Time*, William L. Harper, Robert Stalnaker & Glenn Pearce, editors, 41–55. D. Reidel, Dordrecht.
——. 1970. "Pragmatics." In Stalnaker (1999), 31–46.
——. 1974. "Pragmatic Presuppositions." In Stalnaker (1999), 47–62.
——. 1978. "Assertion." In Stalnaker (1999), 78–95.
——. 1988. "Belief Attribution and Context." In Stalnaker (1999), 150–166.
——. 1999. *Context and Content*. Oxford University Press, Oxford.
Stanley, Jason. 2000. "Context and Logical Form." *Linguistics and Philosophy*, vol. 23 (4): 391–434.
Stanley, Jason & TimothyWilliamson. 2001. "Knowing How." *Journal of Philosophy*, vol. 98 (8): 411–444.
Swanson, Eric. 2006. "Interactions with Context." PhD. dissertation, Department of Linguistics and Philosophy, MIT.
——. 2008. "Modality in Language." *Philosophy Compass*, vol. 3 (6): 1193–1207.
——. 2010. "Propositional Attitudes." Forthcoming in *Semantics: An International Handbook of Natural Language Meaning*, edited by Claudia Maienborn, Klaus von Heusinger, and Paul Portner.
Thau, Michael. 2002. *Consciousness and Cognition*. Oxford University Press, Oxford.
Thomason, Richmond & Francis Pelletier. 2002. "Twenty-Five Years of Linguistics and Philosophy." *Linguistics and Philosophy*, vol. 25 (5–6): 507–529.
Veltman, Frank. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic*, vol. 25 (3).
Wolter, L. 2009. "Demonstratives in Philosophy and Linguistics." *Philosophy Compass*, vol. 4 (3): 451–468.

4.3
THE ROLE OF PSYCHOLOGY

Robert J. Stainton

1 Introduction

Are psychological facts relevant to philosophy of language; and, in particular, does scientific psychology have a legitimate role to play? For example, is it methodologically permissible for philosophers of language to rely upon evidence from neurological development, experiments about processing, brain scans, clinical case histories, longitudinal studies, questionnaires, etc.? If so, why? These two questions are the focus of this survey.

Psychology may seem obviously relevant. I thus begin by introducing arguments against relevance, to motivate the discussion. I will urge that these ultimately fail, and that the appearance of relevance should be taken at face value. In this first section, the focus is on the possible relevance of psychological evidence available to the lay person. Next, I introduce positive arguments for relevance, from examples—shifting the spotlight to recherché evidence drawn from scientific psychology. To foreshadow the main conclusion, psychology, including specifically the methods and results of contemporary cognitive psychology, are relevant because there are connections, both necessary and contingent, between language and the human mind.

2 Two Arguments for Irrelevance and Why They Fail

The negative portion of the paper will focus on two arguments with the same conclusion. Here is the first:

P1: If philosophy of language is not beholden to facts about natural languages, then empirically discovered facts about human psychology are irrelevant to philosophy of language.
P2: Philosophy of language is not beholden to facts about natural languages.
C: Empirically discovered facts about human psychology are irrelevant to philosophy of language.

P1 requires little comment. It is hard to see how human psychology could be probative, if facts overtly about human language are not. What reasons can be given for P2? One might urge that philosophy in its entirety is a priori. Let us ignore this sweeping, radical

idea. More promising is the view that philosophy of language is not "about" natural language. It is about all possible languages, one might say. Or alinguistic propositions. Or how to construct a more perfect language. A thorough response would require an entire chapter. I will content myself with making three brief points. First, the issue is not whether philosophy of language is about actual spoken languages but rather whether it is beholden to empirical results about them. Second, philosophy of language has long paid attention to facts about actual, spoken languages. Historically, philosophers from Plato and Aristotle, through the Cartesians, Hobbes and Locke, were concerned with actual language, and they adduced evidence (as they saw it) from that domain. Similarly, for more recent theorizing: think of recent debates about compositionality, inferentialism in semantics, and the semantics-pragmatics boundary; the question has not been whether a language could have, or should have, such-and-such properties. Nowadays, a great deal of philosophy of language makes use of empirical evidence from theoretical syntax. Third, and finally, I think the tradition has it right: philosophy of language *should be* beholden to facts about human language. Reflection upon why philosophers care about language makes this clear.

Here are examples of philosophical appeals to language, chosen essentially at random. Arguably, the most central debate in 20th-century philosophy of language has been about which expressions are definite descriptions, and what definite descriptions mean. Are descriptions devices of reference, or of quantification? Do sentences containing definite descriptions entail existence, semantically presuppose it, or merely pragmatically suggest it? Philosophers have also argued about whether 'know' is context sensitive in a way that would address skeptical puzzles. They've asked, in the context of analyzing perception, whether 'S saw a green apple' entails 'It seemed to S that something was green', or whether it is even consistent with the latter. Finally, concerned with the nature and existence of moral facts, they have debated whether ethical sentences such as 'Abortion is immoral' have truth conditions. Though I cannot argue the point here, what seems to matter to these larger debates in epistemology, philosophy of mind, and ethics is not merely whether there could be, or should be, sentences (or alinguistic propositions) with the requisite features.

The first argument fails because the antecedent of the major premise is too sweeping to be plausible. The second tries to get by with something weaker. It also adds an "ontological twist":

P3: If philosophy of language is beholden to facts about natural languages, but not to psychological facts, then empirically discovered facts about human psychology are irrelevant to philosophy of language.
P4: Natural languages are not psychological entities.
P5: If natural languages are not psychological entities, then philosophy of language is not beholden to psychological facts.
C: Empirically discovered facts about human psychology are irrelevant to philosophy of language.

To understand the argument, a bit of background will be useful. According to a longstanding philosophical tradition, words are the primary locus of meanings; what they mean are ideas, i.e., items in the mind; and logic describes the mental rules for manipulating ideas. Gottlob Frege revolutionized logic and philosophy of language in the late 19th century by insisting, instead, that theorists look first at complete sentences.

Doing so, we find that meanings are formally structured propositions. For instance, the content of the sentence 'It's not the case that Aristotle is alive' can be decomposed into three quite different semantic parts, none of which has to do with images, sensations, etc. And logic describes not mental operations but truth-theoretic relations among such contents. Thus, as per P4, neither lexical meanings nor semantic relations are ideas or mental processes. Frege also played a key role in the development of artificial logical languages. This provided further support, by affording an alternative model for the ontology of languages: languages are abstract entities, not psychological ones.

Jerrold Katz offers another argument for P4. Consider a perfectly commonsensical distinction between an object O and the human cognizing of O. Mathematics provides paradigm examples: geometry, for instance, is by no means the same as human cognition of geometry. By the same token, Katz (1984: 193) would have us draw "a fundamental distinction between the *knowledge* speakers have of their language and the *languages* that speakers have knowledge of . . . . [T]he subject-matter of linguistics is, in this sense, independent of psychological sciences—just as the subject-matter of logic and mathematics is independent of the sciences concerned with people's logical and mathematical ability". It is only by running these together that languages appear to be ontologically psychological.

Having explained the second argument, let us now evaluate it. I purposely couched P4 and P5 so as to equivocate between two issues. One is whether languages dwell within the mind in the way that dreams, pains, and hallucinations do. This is the issue that Frege and Katz's arguments pertain to. These notwithstanding, many philosophers and linguists follow Noam Chomsky, and insist that languages are mental things in just this sense. (If so, it is plain why psychology would be relevant.) However, we may set that entire debate to one side: The relevance of psychology does not require that languages *be* ideas. Let's focus on another reading of 'psychological entity'—namely, an entity with some kind of necessary connection to (human) psychology. For, the argument's soundness requires that P4 be sustained on that sense as well.

Are languages "psychological entities" in this weaker sense? Languages have both a meaning and sounding side, and they are psychological on both counts. Beginning with the combinatorial aspects of language, the requirements that the rules of syntax be recursive and those of semantics be compositional are driven by psychological considerations, specifically by the empirical fact about human psychology that our minds are finite. That is, natural languages are constrained essentially by our psychology in the sense that, whatever their nature, our limited minds must be able to grasp them. (A similar lesson can be drawn from the fact that no *truly massive* list of sentences could be learned by minds like ours, in the time available.) Turning from combinatorial rules to particular expressions, as Barbara Partee (1979) has stressed, some contents are out-and-out mentalistic: propositional attitude sentences; terms for pain, dreams, mental images, emotions; etc.

What's more, not just exceptionally but in general, linguistic contents are linked necessarily with human psychology. As Chomsky (2000) and Jackendoff (1983, 2002) often insist, many linguistic contents pertain to "objects for us." The things they mean are essentially connected to our properly human interests and points of view. Beginning with two obvious cases, human languages contain words for complex social entities: mortgages, Bollywood, Tuesday. They also include words for fictional and mythical entities. (For example, being an atheist, I think it impossible, wholly independently of human psychology, to give the meanings of: 'venal sin', 'purgatory', and

'transubstantiation'; 'nirvana', 'Navaratri' and 'Ganesh'; 'Qiyamah' and 'Ramadan'; etc.) Other expressions, while picking out ordinary physical objects, nonetheless have contents which are response-dependent. Consider 'cloud', 'tea', 'pet', 'green', 'weapon', and 'bush'. A collection of water droplets is a cloud only if it is perceptible by humans using the naked eye; a pet is an animal that humans treat as such; not all green things share a physically specifiable reflectance property but instead are those items treated "greenly" by the human visual system; and so on. To be clear: the point is not that clouds and pets, mortgages and Bollywood, or even sins, religious holidays and gods, dwell within the human mind. (That is the reading of P4 which we set aside.) Nonetheless, the content of all these words is essentially tied to human psychology. Such are what many, maybe even most, of our quotidian substantives express, and what our demonstratives are prone to pick out in context. Thus P4 properly construed is not so much subject to a few scattered exceptions but is root-and-branch false.

Equally, the "sounding" side of language cannot be individuated independently of human psychology. The philosophical tradition notwithstanding, linguistic expressions are not pairings of acoustic patterns, or waves, with meanings—they are, rather, pairings of something more like clusters of contrastive phonological features with meanings. And the latter are creatures of human psychology. Given space constraints, the details must be set aside. But the point can be made by means of commonsense examples. Think of the various pronunciations of 'cotton'—by a child, a woman, someone with a cold, someone whose larynx has been removed, a second-language learner, or a computer. These are acoustically quite different from one another. What unites such signals is how human minds process them: we hear them as the same. Or again, to the ears of an Anglophone 'van' and 'ban' are different sounds. But for Spanish speakers, these are the same linguistic sound—because 'v/b' is not a meaning-carrying contrast in that language. And English speakers treat the 't' of 'table', 'tea', and 'writer' as the same sound, though their articulation and acoustics are different. Conversely, the very same acoustic wave can realize various linguistic sounds, in different contexts, because of varying semantic import. Finally, acoustic patterns *per se* do not have word boundaries, syllables, vowels, an onset-coda contrast, intonation contours, etc. (See Isac & Reiss 2008: 30–32, 109–114 for introductory discussion.) In short, not just meanings but "linguistic sounds" are necessarily connected to human psychology.

3 A Positive Case for Relevance: Three Examples

I began with two arguments for irrelevance. These helped motivate the questions of whether psychology is relevant to philosophy of language and, if so, why. Roughly speaking, the first very sweeping argument failed because of some of the ends that philosophy of language serves. The second, more circumspect, ontology-based, argument failed because it equivocated on 'psychological entity': reading that phrase such that P4 is *prima facie* plausible, P5 is not; reading it such that P5 is *prima facie* plausible, P4 is not. Those having been rebutted, we may now revert to the null hypothesis, namely that psychological facts are indeed relevant. This speaks to the matter of "whether." With respect to "why," insofar as there are necessary connections between natural languages and human psychology, it would be surprising indeed should evidence about the latter prove irrelevant to the former.

I will now provide a positive case for relevance on the basis of three examples. Notably, they illustrate not just that ordinary, everyday psychology is relevant to

philosophy of language, but that recherché evidence from advanced scientific psychology is. This, of course, further buttresses the positive answer to our first question. The examples also show that, *pace* P5 even on its weak reading, evidential relevance does not require a necessary connection among (a) a background philosophical issue; (b) a corresponding question about actual human language; and (c) evidence from psychology. This yields a second answer to our "why" question, in terms of merely contingent connections.

Proper Names

In laying out the examples, it will be useful to recall a thesis which traces, via W.V.O. Quine, to Pierre Duhem. They note that it is not merely a hypothesis all on its own which yields predictions but rather a hypothesis taken in conjunction with ancillary commitments. Now, hypotheses of philosophical interest are no exception to this rule. In particular, a properly philosophical hypothesis, conjoined with the right ancillary hypotheses, can yield a prediction which is verified or falsified by scientific psychology.

For instance, to introduce our first hypothesis H, it has been suggested that *names share the content of descriptions of persons*. There are clear philosophical motivations for H, both metaphysical and epistemological. Very roughly, if proper names are synonymous with descriptions, then nonreferring ones can have meaning without undue ontological commitment; they won't embed *salva veritate* in certain modal contexts; and personal identity will presumably depend upon the retention of the descriptive properties. What's more, turning to epistemology, deploying a name can afford knowledge of the person absent acquaintance with its bearer; and propositional attitudes will not pertain to "persons themselves," but rather to descriptive contents. (It is precisely because H is motivated philosophically in these ways, that it is a hypothesis within philosophy of language.) Now, H itself is not a claim about human psychology. It may even be, *pace* the discussion above, that the truth or falsity of H is metaphysically independent of human psychology. Nonetheless, the methods and results of scientific psychology are relevant to its evaluation. That is because H can be (dis)confirmed in conjunction with ancillary commitments A, including postulated laws.

As Valentine et al. (1996) explain at length, proper names are psychologically special. Processing them is notoriously slow; they are especially vulnerable to brain damage; and, as experience of "tip of the tongue" moments makes intuitively clear, names are hard to remember. (Experiments confirm this: when subjects are presented with unfamiliar faces, and taught various facts about the people shown, including their names, the latter prove significantly harder to recall. See Cohen and Faulkner 1986.) It seems, and this is another element of A, that the best explanation of this "specialness," is some kind of psychological law pertaining to the content of names; that, rather than their form, seems to be why they pattern this way. H and A together yield a prediction P, namely that a synonymous description of a person will exhibit the same psychological profile as the corresponding name. However, as experimental, clinical, and other work shows, this is not the case. Thus, a hypothesis in philosophy of language is called into question by evidence from sophisticated scientific psychology. (See Segal 2001 for detailed and careful discussion.)

Knowledge Attributions

There has been much philosophical debate about the conditions under which knowledge attributions are strictly speaking true. One motivation for the debate is a Moore-style argument against external world skepticism. The idea goes that, as ordinary speakers use the sentence 'Rob knows that Hitler is dead', the sentence is true; and, so used, the truth of this sentence entails that Rob has knowledge of the external world. So, such knowledge exists. A natural reply to this argument is to insist that, *appearances notwithstanding, the sentence is false*. This is H. Of course, one cannot just stop there: one must explain away the strong intuition that the sentence is true. The reply thus continues: 'Rob knows that Hitler is dead' strikes people as true because they confuse what it literally says with things that its usage merely conversationally implicates, for example, that Rob's belief is warranted, that one can safely proceed on his say-so, etc. This explanation of the appearances is A. Again, H is not a claim about the human mind. Maybe its truth is not even conceptually, logically, or ontologically connected to cognitive scientific facts. Nonetheless, H conjoined with A yields a prediction: that people who cannot grasp conversational implicatures will reject everyday knowledge attributions as false. This prediction, P, is disconfirmed by evidence from speakers with Autism Spectrum Disorders (ASDs): despite exhibiting good formal language abilities, people with ASDs suffer from serious pragmatic impairments, including in particular failing to notice conversational implicatures. Yet, *pace* P, they do not present as external world skeptics. (Other deficits should yield related (dis)confirmation, by the way—specifically Semantic-Pragmatic Disorder and Right Hemisphere Dysfunction. For a useful survey, see Asp and de Villiers (2010) and Cummings (2009). Incidentally, ASDs afford another interesting evidential link between clinical psychology and philosophy of language. Linguistically able people with ASDs famously have trouble attributing complex mental states to others. This would seem to call into question the Gricean (1957) idea that meaning something, for example making a statement, of necessity involves higher-order intentions. See Andrews 2002 and Gluer and Pagin 2003 for discussion.

The import of the foregoing examples is not, of course, that considerations from scientific psychology have established that names are directly referential rather than descriptive, or that external world skepticism is false. The point is that, even barring necessary connections, Duhem's Thesis yields links between philosophical hypotheses involving language and evidence from cognitive psychology: Because of diverse and unimagined contingent (including nomic) connections, "properly philosophical" hypotheses about language give rise, in unpredictable ways, to empirical predictions about the mind.

Assertion

A final example illustrates just how far removed the background philosophical issues can be from the curious psychological findings. In the context of defending metaphysical antirealism, it has been proposed that truth be explicated in terms of warranted assertion. Doing so, of course, means that assertion itself cannot be analyzed in terms of truth: that would be circular. A promising way out is to *explain the speech act of assertion socially/formally, specifically in terms of the use of declarative sentences in conventionally specified conditions* (Dummett 1973). This is H. Now, if H holds, then, appearances notwithstanding, subsentential words and phrases cannot be used to assert. Instead, for

instance, when someone displays a letter and says of it 'From Chomsky', what they produce must be an elliptical declarative sentence. This attempt to explain away a seeming counterexample to H yields P: someone who cannot engage in sentential ellipsis cannot make such an assertion. But P is open to psychological (dis)confirmation. In particular, people who cannot yet, or can no longer, form or comprehend elliptical sentences, can nonetheless make and understand assertions with bare words and phrases. (See Stainton 2006 for extended discussion.) Granted, it seems outlandish that antirealism and the nature of truth could be tied to aphasias and child language. And yet, the former have been connected, by hypothesis, to the speech act of asserting; and thence to a formative of a certain kind, the declarative; and this philosophical line of thought necessitates endorsing a linguistic conjecture about ellipsis; and so, by a long and winding road, the whole account becomes subject to arcane psychological evidence.

4 A Parting Shot: Irrelevance In Practice and In Principle

My contention in this chapter, with respect to our first question, has been that psychology, including advanced scientific psychology, does have a legitimate evidential role in practice. With respect to the second question, namely why this should be so, the main lesson is this: the actual nature of human language is relevant to certain properly philosophical questions; and, because of both contingent and necessary connections, human psychology is relevant to the actual nature of human language; that is why, by the transitivity of 'relevant to', psychology is relevant to philosophy of language.

I have not restricted myself to the issue of whether such evidence is irrelevant *in principle*. Surprisingly, certain philosophers have rejected the latter view as well. So, I will end with a word about it.

It does remain epistemically possible that philosophy of language is disconnected from psychology in general, and scientific psychology in particular. Hence it could still turn out, in the long run, that MRIs, error rates, etc., are of no use. Nonetheless, as Fodor (1981) has stressed, as long as we do not know what is in fact connected to what, and do not know in particular what linguistic facts supervene on, we should not rule out psychological evidence a priori. Consider again in this light the three examples above. What rebuttals are rationally permissible? It would be natural to insist that the alleged psychological data are incorrect, or at least oversimplified. Or again, one could grant that the data are accurate but reply that it can be accommodated to, or even supports, the philosophical account at issue. For example, one might scrutinize more carefully the speech of people with ASDs, and counter that they actually do use 'know' as the anti-Moorean predicts. Maybe names and descriptions do indeed share the same meaning but behave differently with respect to processing for some other reason. And maybe, as Stanley (2000) has argued, some apparent uses of words and phrases are elliptical after all, while others simply do not amount to assertions. All these rebuttals are appropriate. However, they involve bringing in more psychological evidence, not less. In contrast, what should emphatically not be accepted is simple dogmatic defiance: "That's all very fascinating. But I refuse even to offer a response. Those are psychological findings, and I already know full well, on the basis of metaphilosophy and the metaphysics of language(s), that such results *cannot* pertain to the philosophy of language."

This alone shows that, even if at the idealized end of inquiry there turned out to be no connection between philosophy of language and psychology, nonetheless the latter should not now be dismissed out of hand. Put otherwise, only an omniscient creature

could have a handle on what "the considerations in principle relevant to philosophy of language" are. Not being such creatures, we are well advised to pursue evidence which seems relevant—even when it comes from psychology.

References

Andrews, Kristin (2002). "Interpreting Autism". *Philosophical Psychology* 15(3): 317–332.

Asp, Elisa and J. de Villiers (2010). *When Language Breaks Down: Analysing Discourse in Clinical Contexts*. London: Routledge.

Chomsky, Noam (2000). *New Horizons in the Study of Language and Mind*. Oxford: Cambridge University Press.

Cohen, G. and D. Faulkner (1986). "Memory for Proper Names." *British Journal of Developmental Psychology* 4: 187–197.

Cummings, Louise (2009). *Clinical Pragmatics*. Cambridge: Cambridge University Press.

De Villiers, Jessica, Robert J. Stainton and Peter Szatmari (2007). "Pragmatic Abilities is Autism Spectrum Disorder". In P.A. French and H.K. Wettstein (eds.) *Philosophy and the Empirical*. Volume XXXI of *Midwest Studies in Philosophy*. Oxford: Blackwell, pp. 292–317.

Dummett, Michael (1973). *Frege: Philosophy of Language*. London: Duckworth.

Fodor, Jerry (1981). "Some Notes on What Linguistics Is About". In N. Block (ed.) *Readings in Philosophy of Psychology, Volume 2*. Cambridge: Harvard University Press, pp. 197–207. Reprinted in J.J. Katz (ed.)(1985). *The Philosophy of Linguistics*. Oxford: Oxford University Press, pp. 146–160.

Gluer, Kathrin and Peter Pagin (2003) "Meaning Theory and Autistic Speakers." *Mind and Language* 18(1): 23–51.

Grice, H. Paul (1957). "Meaning." *Philosophical Review* 66: 377–388.

Isac, Daniela and Charles Reiss (2008). *I-Language*. Oxford: Oxford University Press.

Jackendoff, Ray (2002). *Foundations of Language*. Oxford: Oxford University Press.

—— (1983). *Semantics and Cognition*. Cambridge, MA: The MIT Press.

Katz, Jerrold (1984). "An Outline of Platonist Grammar." In T.G. Bever, J.M. Carroll, and L.A. Miller (eds.) *Talking Minds*. Cambridge: The MIT Press, pp. 1–33. Reprinted in J.J. Katz (ed.)(1985). *The Philosophy of Linguistics*. Oxford: Oxford University Press, pp. 173–203

Partee, Barbara (1979). "Semantics—Mathematics or Psychology?" In R. Bauerle, U. Egli, and A. von Stechow (eds.) *Semantics from Different Points of View*. Berlin: Springer, pp. 1–14.

Segal, Gabriel (2001). "Two Theories of Names." *Mind and Language* 16(5): 547–563.

Stainton, Robert J. (2006). *Words and Thoughts: Subsentences, Ellipsis and the Philosophy of Language*. Oxford: Oxford University Press.

Stanley, Jason (2000). "Context and Logical Form." *Linguistics and Philosophy* 23: 391–434.

Valentine, Tim, Tim Brennen, and Serge Bredart (1996). *The Cognitive Psychology of Proper Names*. London: Routledge.

4.4
THE ROLE OF MATHEMATICAL METHODS

Lawrence S. Moss

1 On the Contribution of Mathematics to the Study of Language

Mathematical models have contributed to our understanding of linguistic phenomena. This chapter makes the case for this, and at the same time it tries to do so in a way that highlights problematic points. I would like to start with a few theses about the uses of mathematics in linguistics:

(1) Applying mathematical ideas has changed the kinds of questions we ask about language.
(2) Applying mathematical ideas has enabled us to use metaphors from mathematics.
(3) Applying mathematical ideas has facilitated explorations of structure.
(4) In order to get computationally sensible models of language, it is necessary to apply a range of mathematical techniques and ideas.

I am going to explore all of these points, in a general way in this section and then in expanded case studies in the remainder of the chapter. The overall point is that *mathematical modeling* in the study of language should be seen in the same light as mathematical modeling in other areas, say physics or sociology; it is a useful tool, perhaps not the only useful tool, but a very useful one nonetheless: At the same time, it must be admitted that not all applications of mathematics are successful. Finally, it is a difficult matter in the philosophy of mathematics to account for the very usefulness of mathematics in the first place. Why should it be that a discipline which traffics in abstractions be so connected to the concrete? How is it that facts whose truth seems necessary come to be connected to a world so contingent? We shall not take up these issues here; they are far beyond our scope. But they form a backdrop to the issues that is worth keeping in mind.

The first point about the questions we ask is partly supported by a quote from Noam Chomsky's book *Syntactic Structures*, arguably among the most influential books in the history of linguistics:

> Precisely constructed models for linguistic structure can play an important role, both positive and negative, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by *ad hoc* adjustments or loose formulation.

We take a "precisely constructed model" to be a mathematical model, even if no mathematics is used. What Chomsky means here by a "negative" role is that even a theory which leads to bad predictions might help people construct a better theory. It is more interesting that the "positive role" that he mentions is not just that one can use a model to solve a problem, it is that the model might enable one to find "solutions for many problems other than those for which it [the theory] was explicitly designed."

My second point, on metaphors, is a point worth making on practically all uses of mathematics. One rarely uses the deepest results of any topic in mathematics a instead it is the; transfer of ideas from the abstract world of mathematics to the applied area. For example, when one sees applications of dynamical systems in cognitive science (see, e.g., Port and van Gelder (1995)), it is largely the metaphors related to concepts like *attractor* connected to the long-term qualitative behavior of dynamical systems. One rarely sees applications of the best-known theorems in the subject. For linguistics, this point is obscured when we turn to most uses of mathematics in areas like syntax, for there the mathematics was developed specifically with the application, or something like it, in mind. However, it does hold for uses of logic. I cannot think of any specific results from logic that are used in any area of linguistics. To be sure, one finds uses of results such as the Cut Elimination Theorem for various logical systems used in connection with grammar formalisms, to mention just one example. But nothing is used in a widespread way. On the other hand, some of the underlying ideas of modern logic such as the *syntax/semantics distinction* and the notion of *formal derivation* are as important in formal semantics as they are in logic. This point on metaphors will be discussed a bit more when we turn to the examples of phylogenetic reconstruction.

Point (3) is motivated by the fact that language comes to us without any evident structure. It is up to theoreticians to propose whatever structures they think are useful, and these include structures related to all areas of language from the phonological level to the level involving the evolution of human language. Mathematical models are the primary way that scientists in any field get at structure.

We now take up point (4) on the use of mathematical ideas in computational settings involving language. I am thinking of *natural language processing* (NLP) and *computational linguistics* (CL). Among the fields of mathematics that have been used in connection with NLP and CL are: automata theory (many of the tools in actual use come from this area), linear algebra (in connection with *latent semantic analysis*, to mention just one illustration), algorithms and computational complexity theory, and probability theory and statistics. It is hard to imagine much serious work in NLP and CL without these mathematical tools. (Incidentally, this point might be what Chomsky had in mind in the quote above by

using a model to find solutions to problems other than the one in which one was originally interested. His pioneering results on regular and context-free grammars are not muchused in syntax; they are practically forgotten there. But they are of great use in NLP and CL.)

Going further than linguistics, let me make a point related to the uses of mathematics in the broader field of cognitive science. When one thinks of mathematics as an applied discipline, the leading application areas would probably be physics and engineering. However, this is not the only way of motivating the subject: one could instead think about the cognitive sciences, including linguistics, psychology, artificial intelligence, and the like. A very large number of mathematical disciplines could instead be motivated by their application in these fields. Indeed, one could well imagine that someday these areas of applied mathematics might be seen as the centerpieces of the subject.

1.1 Objections to Mathematics

Before taking up objections to the theme of this chapter, we should mention a few points which we are *not* claiming.

(5) We are not claiming that *all* of the applications of mathematics in linguistics are correct or even interesting.
(6) We are not claiming that mathematical tools are the only valid ones for the study of language.

Regarding the first point, it is actually difficult from our vantage point concerning mathematical modeling to make the claim that one or another model is *incorrect*; the strongest assertions we can make are that the end results of the modeling are empirically false, or that they conflict with what "everyone in the field knows," or that they lead to uninteresting questions. Even with this in mind, we do not claim that all uses of mathematical ideas in any given field are useful or interesting. With regard to contemporary linguistics, this is trivial point: on any given debate, probably proponents of all sides marshall arguments and theories that are on our view applications of mathematical methods.

It might be good at this point to mention some studies related to language that do *not* use mathematical methods. One example could be comparative philology, the study of texts to infer relations between languages (exemplified by the deciphering of the Rosetta Stone by Jean-François Champollion in 1922). Another, closer to this volume, could be areas connected to the philosophy of language which have no trace of mathematics.

Here are a few objections to the use of mathematics in connection with the study of language:

(7) In mathematics, everything is either true or false, black or white. Thus mathematics would never be able to accommodate the kinds of *degrees of truth* that we see in language: words partly mispronounced, sentences partly grammatical, ideas that do not reflect anything with perfect clarity.
(8) Language is a psychological and social phenomenon. Therefore, the only legitimate form of research on language is via psychology and sociology, or other fields that place human phenomena at the forefront. Using mathematics leads to a sterile science that ignores what is most interesting about language.

(9) There is nothing wrong with the use of mathematics in connection with language. In principle, it could be a good thing. However, nothing so far has been done that is of lasting value to linguistics.

One reply to objection (7) is that mathematics does not commit us to the view that the objects it models bear the same kind of clear-cut distinctions that one does indeed find inside of mathematics. To take one example, consider probability and its applications. The standard *theory* of probability proceeds the same way as any field of mathematics: There are definitions and axioms, and statements that either follow from the axioms or which do not. People who work on the subject almost never disagree on claims about what is true in the subject, and what is false. (Of course, they may well disagree on the intended applications of the theory.) But all of the uses of probability theory are specifically about phenomena that are not clear cut, that exhibit randomness, chance, and unpredictability. Turning specifically to language, we do indeed find many applications of probability theory and statistics. In fact, there are many uses of probability in connection with syntax in natural language processing.

Nevertheless, there is an aspect of (7) that does seem apt. Many of the items in this handbook are applications of *logic*. (Some of them are actually applications *to* logic, but we set these aside for this discussion.) The predominant methodology in logic is indeed that sentences are true or false. (There is a field called *paraconsistent logic* which holds that sentences can be *both* true and false. The view that sentences might be neither true nor false is fairly commonplace.) And so logic might be the wrong tool for doing certain kinds of modeling. However, using *mathematics* does not commit us to using *modern logic* (or any particular branch of mathematics), so criticism (7) seems ill-placed. To state our reply again, in this chapter the use of *mathematical methods* is understood very broadly. It encompasses neural networks for example: they enable one to use mathematical metaphors and ideas just as surely as other forms of modeling.

Objection (8) seems valid at first glance, but as soon as one looks into the literature in the social sciences, one indeed does find applications of mathematical ideas and results. So this objection would seem to fall flat. For example, serious work on human memory makes use of mathematical models. (Alternatively, the objection to mathematics in linguistics would have to be widened to an objection to many other applications of mathematics. With a wider circle of applications, it would seem to be a harder and harder case to make.)

I mention the last objection mostly because it is forcefully made by Noam Chomsky and yet contrasts sharply with his 1957 position. In 1993 he gave the Patten Lectures at Indiana University. At a lecture attended by linguists, logicians, and philosophers, he was asked what results of mathematics had the most implications for linguistics. His answer was essentially that none of the applications had any real value, with one possible exception (the Chomsky-Schützenberger Theorem). Further, he took a position that I have outlined as objection (9). In a sense, this chapter as a whole is a kind of answer to this objection.

2 Examples and Case Studies

The remainder of this chapter discusses two applications of mathematics to linguistics which either expand on what we have already mentioned or which present issues that are worthy of further discussion. If we had more space, we would have gone into details

on the application of game theory to pragmatics, since it would have fit the topic of the handbook. For a collection of papers on this topic, see Benz et al. (2006). We might also have discussed generalized quantifiers, and also the mathematics of learnability.

2.1 Phylogenetic Estimation in Historical Linguistics

One stunning application of mathematics to the study of language pertains to *historical linguistics* and the *reconstruction or estimation* of the phylogenetic tree of languages. As the name suggests, there is a parallel with biology. Consider the following trees:

The one on the left is intended as a representation of the evolution of a family of languages, with the edges representing descent in a very long time period. On the right we have a similar tree for part of the primate family. Please do not take the content of these trees seriously; we are only presenting them as examples. The trees in both historical linguistics and in evolutionary biology share some key features: They represent long-term change, they are partly derived from historical evidence (texts or bones), but the evidence is scant, and they are frequently the topics of debate. The same mathematics is used in both cases; it is mainly probability and algorithms. In fact, the same people work on both topics.

From the point of view of this Chapter, the interesting question concerns the role of the technical work itself, and so we return to the four points mentioned at the beginning of this chapter. Work on phylogentic reconstruction may well change the questions historical linguists ask. In broad terms, the way that one uses linguistic data to reconstruct trees is to compare features of living languages such as cognate words (e.g., English *one* and Spanish *uno*) or parallel syntactic structures (whether the language mainly has verbs at the end or not). The actual choice of the features is controversial, and so the results of work in the area could inform the practice in historical linguists. Further, the way one uses the data in reconstructing or estimating the trees is also controversial. That is, there is a host of algorithmic options, and very little in the way of actual records to know what counts as a correct tree. One can propose a computer simulation of the evolution and then run algorithms on this. But in the case of language, we again have the problem of justifying the use of the simulations. In any case, the actual practice in this area might well contribute to a clarification of the questions that people in the field ask. I take this to be the most interesting connection to the topics of this chapter. In addition, we might well find metaphors used in computational biology finding their way into historical linguistics.

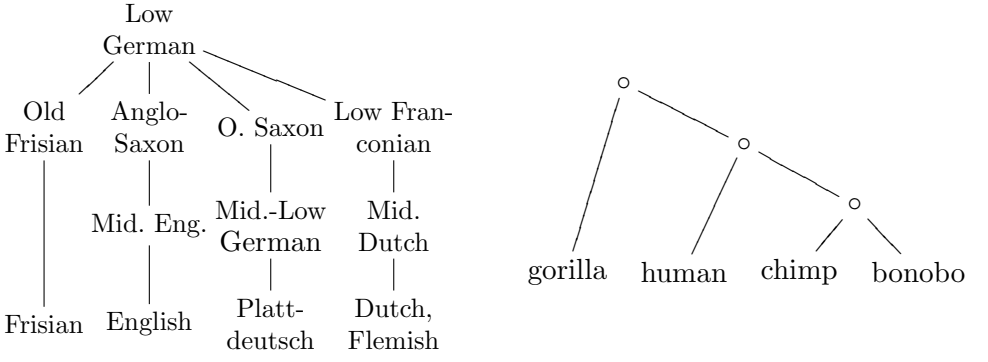


Figure 4.4.1

Another interesting point for us is that, at least so far, the work in this area is not universally accepted:

> Over the last 10 or more years, there has been a tremendous increase in the use of computational techniques (many of which come directly from biology) for estimating evolutionary histories of languages. Despite the enthusiasm with which the popular press has received these studies, it is probably fair to say that much of the community in historical linguistics has been skeptical of the claims made by these studies, and perhaps even dubious about the potential for such approaches to be of use.

This is from an overview tutorial of current work on this topic; see Nichols and Warnow (2008). The question, then, is what all the work comes to. It is possible to use phylogenetic reconstruction to address questions of historical linguistics. For example, Warnow et al. (1996) consider long-standing questions concerning Indo-European languages, such as whether the Italic and Celtic branches should be siblings. But it is hard to make a definitive assessment of the meaning of the results. Overall, this example of mathematical methods in the study of language underscores both the usefulness of those methods, and also some of their troublesome aspects.

2.2 Infinity and Recursion in Language

Our most extended discussion centers on the relation of language to *infinity*. It is often claimed that every human language is in a precise sense *infinite*, and that this point is important. At the same time, the matter is not without considerable controversy. The touchstone of the controversy is not so much *infinity* per se but its cousin, *recursion*. Writing in a popular journal *Science*, Hauser et al. (2002) state:

> (10) All approaches agree that a core property of FLN [the faculty of language in the narrow sense] is recursion . . . . FLN takes a finite set of elements and yields a potentially infinite array of discrete expressions. This capacity of FLN yields discrete infinity (a property that also characterizes the natural numbers) . . . . It has been recognized for thousands of years that language is, fundamentally, a system of sound-meaning connections; the potential infiniteness of this system has been explicitly recognized by Galileo, Descartes, and the 17th-century "philosophical grammarians" and their successors, notably von Humboldt.

A *set* is a definite collection of objects. These are usually mathematical objects, but it is crucial in this discussion to consider also sets consisting of words, sentences, and the like. As a mathematical preliminary, the set N of natural numbers is the set

$$\{0, 1, 2, \ldots, n, \ldots\}$$

A set S is *finite* if there is some natural number n such that the elements of S may be matched with the set $\{1, 2, \ldots, n\}$ in a manner which is *one-to-one* (no repeats in the matching) and *onto* (every number in $\{1, 2, \ldots, n\}$ is matched with some $s \in S$. A set is *infinite* if it is not finite.

For example, the set {△, o} is finite, and the set *E* of even numbers is infinite. Also, it is a standard fact of set theory that *N* is not a finite set. (This is not entirely obvious, since our definition of *finite* was made independently.) Further, no set which has a matching with *N* is finite, either.

The main claim under discussion here is

(11) The set of sentences in every natural language is an infinite set.

We shall say later *why* one would want to make this claim (11) and also *how* one argues for or against it. But before we turn to those points, it should be noted that the claim that natural languages are infinite is not to be found in other disciplines besides modern linguistics. It seems to be a by-product of asking mathematical questions about language.

The main reasons for (11) have to do with what are sometimes called *recursive constructions* in language. Among these are boolean compounding, intensification of adjectives using *very*, iterations of possessives, and iteration of propositional attitudes. To be specific, here is one concrete argument for (11) in the special case of English.

Let

GP = {my great^n grandparents : $n \in N$}.

In a little more detail, $(great)^n$ denotes the result of writing down the word *great n* times in a row. Thus $great^1$ = *great*, $great^2$ = *great great*, etc. So the first item in *GP is my grandparents*, since with $n = 0$ we simply drop the word *great*. The second item in *GP is my great grandparents*, etc.

We now have a matching between *N* and *GP*:

(12)

N	*GP*
0	my grandparents
1	my great grandparents
2	my great great grandparents
. . .	. . .
n	my (great)^n grandparents
. . .	. . .

The reason that (12) describes a matching is that if *j* and *k* are different numbers, then *my* $(great)^j$ *grandparents* and *my* $(great)^k$ *grandparents* have different numbers of words and hence are different strings of words. It is clear that each element of *GP* is a noun phrase (NP) of English. Thus the collection of English NPs is infinite. Similarly, we can find a collection of English sentences which is infinite: just take *GP* and append *died* at the end.

However, the *subtext* for claiming (11) is that (11) provides an argument for the *generative grammars*. A generative grammar is a kind of mathematical object which embodies the idea of building linguistic structures from the "bottom up" rather than the "top down." That is, by a generative grammar one means one of a large number of alternative definitions; what exactly counts as a generative grammar is, I think, up to debate, and I am very sure that there are going to be many borderline cases. What the various flavors of generative grammars have in common is that they employ or reflect a metaphor of

building structure starting with basic, unanalyzed forms, and following this they expand on these by using rules which build complex forms from simpler ones.

An interesting point here is that made in Tiede and Stout (2010):

> Hauser et al. (2002) claim that a core property of the human language faculty is recursion and that this property "yields discrete infinity" of natural languages. On the other hand, recursion is often motivated by the observation that there are infinitely many sentences that should be generated by a finite number of rules. It should be obvious that one cannot pursue both arguments simultaneously, on pain of circularity.

The message of the quote is that it is not always clear whether one is using infinity to justify recursion or recursion to justify infinity. Mathematically, the latter move is legitimate while the former is not. And one reading of arguments in support of recursion in language is that they are disguised forms of arguments for the generative approach to grammar. For this reason, writers such as Pullum and Scholz (2010) who are ill-disposed toward generative grammars have cast a suspicious eye on the argument for infinity based on recursion, and even on (11) as it applies to English. However, we maintain that the argument above for (11) is valid. The only assumption that should be debated is whether all rows of the chart in (12) are indeed NPs of English. (Of course, this is tantamount to the infinitude claim concerning English in the first place.)

Now what are the arguments against (11)? One claim is that the collection of sentences in a natural language like English is not a *set* at all. That is, it is not a "definite collection of objects." Whether a given string S of words is a sentence might not be a clear-cut matter. One person might hold S to be a sentence, and another to deny this. Even worse, the same person might hold differing opinions at differing times (or even at the same time).

But if a language is not a definite collection, is there any way to think about it mathematically? There are indeed many such ways, and one would be as a *probabilistic language*. This would be a function ℓ from strings on some alphabet to the unit interval $[0, 1]$. In the case at hand, the "alphabet" would be the words of English (not the letters). For example, one might interpret $\ell(S)$ as the probability that a randomly selected speaker of English assents to the assertion that S is a sentence of the language.

Returning to (11), one could object to our argument on the grounds that for some large n, *my* $(great)^n$ *grandparents* is not a noun phrase of English: perhaps when n is larger than the number of particles in the universe, for example. So the matter of infinity and recursion then hinges on a deeper matter, whether we model English as an idealized collection in the first place. For more on this point, see Tiede and Stout (2010).

But what about other languages? Up until now, all of our work has been on the special case of the claim in (11) that the set of sentences of *English* is infinite. What about other languages? What about the stronger claim that (11) makes? For the most part, one senses that there is a consensus that the argument should apply to all languages. However, there is an argument that this is false. The most recent version of this comes from work of Daniel L. Everett on the language Pirahã, spoken in northwestern Brazil. In papers and also in publications such as Everett (2007), he claims that there is *no recursion* in the syntax of Pirahã:

> So in the case of Pirahã, the language I've worked with the longest of the 24 languages I've worked with in the Amazon, for about 30 years, Pirahã doesn't

> have expressions like "John's brother's house". You can say "John's house", you can say "John's brother", but if you want to say "John's brother's house", you have to say "John has a brother. This brother has a house". They have to say it in separate sentences.

Indeed, this casts doubt on (11) in the form that we have stated it, and hence on the quote in (10). That is, if recursion is not a feature of all human languages, it is then not a universal of language the way (10) claims, and so what is the point of making it the centerpiece of grammar?

Our purpose here has been to discuss the matter both for its own sake and also as an illustration of the main points in the Chapter, mathematical methods in the study of language. Our sources for this discussion are Pullum and Scholz (2010), Tiede and Stout (2010), and Keenan and Moss (n.d.).

2.3 Refinement: Are Natural Languages Regular?

A further refinement of the question concerning the infinity of language concerns the place of a natural language in the *Chomsky hierarchy*. To state the issue, we need some definitions from *mathematical linguistics*. We need the definition of *regular languages*, and we also need to apply the *Pumping Lemma* for them. One can find these topics in any book on formal language theory or mathematical linguistics. But even without these, we can discuss the matter informally. From a mathematical point of view, not all infinite sets are created equal. There are many ways to define scales which measure *complexity*, for example. At the very bottom end of one such scale, we have what are called the *regular languages*. A *regular expression on the words of English* is an expression such as

> my great^{+} grandparents
> the(boy + girl) swam
> (the (friend + enemy) of)$^{+}$ the President died

Each of these has a *denotation*, a set of sequences of words of English. The denotation of the first is listed in (12). The denotation of the second is the set with just two elements, the sequences *the boy swam* and *the girl swam*. In effect, the + means "make a choice." The + means "repeat one or more of times." So the denotation of the third set is an infinite set, including the following:

> the friend of the President died
> the enemy of the President died
> the friend of the enemy of the President died
> the friend of the enemy of the enemy of the President died

A *regular language* is a set of sequences of words which happens to be the denotation of a regular expression on the words of English. We want to present an argument that English is not a regular language. That is the set of English sentences cannot be the denotation of any regular expression, no matter how long, complicated, or convoluted.

Here is the proof: suppose toward a contradiction that English were regular. It is a standard fact that the intersection of two regular languages is regular. This is not at all obvious from our definition. But granting this, consider the language *L* defined by

L = English $\cap$ {the, boy, girl, saw, ran}$^+$

From the point about intersections of languages, this language L must also be regular.

Here are some sentences in L:

the girl ran
the girl the girl saw ran
the girl the girl the girl saw saw ran
the girl the girl the girl the girl saw saw saw ran

We know that nobody can possibly understand this last sentence, but there are reasons why linguists want to take it to be part of *ideal English*. These are the same reasons that one would want to take all of the expressions in (12) to be English NPs.

The main points about L are that

(13) No sentence in L has two "the"s in a row.
(14) No sentence in L has two nouns in a row.
(15) Every noun phrase in every sentence in L has exactly as many nouns as verbs.

The first assertion is a general fact about English. The other two are justified by examining L. If we assume that *ran* cannot take a direct object, then the sentences above are a full listing of L. If we use *ran* transitively, then we get sentences like *the girl saw the girl the girl saw*, with *the girl the girl saw* as an object noun phrase. But these still conform to the last point. Consider a long sentence S_n in English of the form

(the girl)$^{n+1}$ saw^n ran,

with n to be determined shortly. Our sentence S_n belongs to L, and S_n must have one more "girl" than "see".

At this point, we wish to use a famous result about regular languages called the *Pumping Lemma*. It would take us too far afield to state it formally, but the gist of it is that for any given regular language, there is a number n so that if we have a sequence that belongs to the language whose length is at least n, then there must be some repeatable piece of the first n words. Informally, any regular expression whose denotation is the given language would have to involve an expression with the superscript +, and then we get some repetitive structure. More formally, the lemma tells us that we may (a) find some number n; (b) to decompose the sentence S_n with this n as xwy for some non-empty subword w of S included in the first n words of S_n; (c) "pump" $S_n = xwy$ by repeating w, to get xw^2y; and (d) do all this in such a way that xw^2y is also in the set L. The Pumping Lemma does it all. Getting back to our sentence S_n and the decomposition xwy, let us ask what w can possibly be. It must be (*the girl*)^k for some k, or else (*the girl*)^k *the*. But this string xw^2y would not belong to L: its subject noun phrase would have more nouns than verbs. Since xw^2y is not in L, we have a contradiction. This means that English cannot possibly be regular.

For more on this topic and related ones, see the collection Savitch et al. (1987).

As we end let us reflect on the interaction of mathematics and linguistics in it, as a reprise of our original points in Section 1.

The argument that we have seen can be a motivation to look at an important grammatical construction, *center-embedded relative clauses*. This hints at the contribution of mathematical methods in syntax, a point which seems to me to be clear. There are also many metaphors that come from mathematics and computer science to the study of syntax. The notion of *structure* is especially interesting and problematic in this discussion. In the first place, the general feeling is that the formalism of regular languages does not get at the kind of structure that linguists traditionally favor. So the argument we have seen may be used to motivate looking beyond the regular languages. This was the original purpose in the argument. In contrast, opinions are somewhat mixed about whether natural language "should" be regular in the first place. My sense is that the majority feel that constructions like center embedding (and the argument we have seen) indeed show that English is not regular. For other languages, there are even arguments to the effect that they should not be *context-free*. A few people reject the argument and feel that English is either outright regular, or "fairly close." Turning to our last general point, although the regular languages are not so much used in contemporary linguistics, they form a very useful tool in computational linguistics. All in all, the points which we made in this section are typical points about the useful and sometimes problematic relation of mathematics and language.

Acknowledgments

My thanks to the editors of this volume for their patience. I am also grateful to Robert Rose and Hans-Jörg Tiede for comments and discussion. All the remaining deficiencies are entirely my own responsibility.

References

Benz, A., Jäger, G. and van Rooij, R., eds. (2006), *Game Theory and Pragmatics*, Palgrave Macmillan.

Everett, D. L. (2007), Recursion and human thought: why the Pirahã don't have numbers. See http://www.edge.org/3rd_culture/everett07/everett07_index.html

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002), "The faculty of language: What is it, who has it, and how did it evolve?," *Science* 298(5598), 1569–1579.

Keenan, E. L. and Moss, L. S. (n.d.), *Mathematical Structure in Language*. Unpublished ms., UCLA.

Nichols, J. and Warnow, T. (2008), "Tutorial on computational linguistic phylogeny", *Language and Linguistics Compass* 2(5), 760–820.

Port, R. and van Gelder, T. (1995), *Mind as Motion: Explorations in the Dynamics of Cognition*, Bradford Books/MIT Press.

Pullum, G. K. and Scholz, B. C. (2010), Recursion and the infinitude claim, *in* H. van der Hulst, ed., *Recursion in Human Language*, Vol. 104, Mouton de Gruyter, Berlin, pp. 113–138.

Savitch, W. J., Bach, E., Marsh, W. and Safran-Naveh, G., eds. (1987), *The Formal Complexity of Natural Langugage*, D. Reidel, Dordrecht.

Tiede, H.-J. and Stout, L. N. (2010), Recursion, infinity, and modeling, *in* H. van der Hulst, ed., *Recursion in Human Language*, Vol. 104, Mouton de Gruyter, Berlin, pp. 1–12.

Warnow, T., Ringe, D. and Taylor, A. (1996), Reconstructing the evolutionary history of natural languages, in *Proceedings of ACM–SIAM Symposium on Discrete Algorithms (SODA)*, Association for Computing Machinery and the Society of Industrial and Applied Mathematics, pp. 314–322.

4.5
THE ROLE OF ARTIFICIAL LANGUAGES

Martin Stokhof

1 Introduction

When one looks into the role of artificial languages in philosophy of language it seems appropriate to start with making a distinction between philosophy of language proper and formal semantics of natural language. Although the distinction between the two disciplines may not always be easy to make since there arguably exist substantial historical and systematic relationships between the two, it nevertheless pays to keep the two apart, at least initially, since the motivation commonly given for the use of artificial languages in philosophy of language is often rather different from the one that drives the use of such languages in semantics.

Of course, this difference in motivation should not blind us for the commonalities that exists between the two disciplines. Philosophy of language and formal semantics have a common history, and arguably also share some of their substance. Philosophy of language is by and large an outgrowth of work in the analytical tradition in philosophy in the first half of the twentieth century. Both ordinary language philosophy, with its emphasis on the description of actual language use, as well as the more logic-oriented and formally inclined school of logical positivism contributed to the definition of philosophy of language as a separate philosophical discipline, with its own set of problems and methods to solve them. Another major contributor to the establishment of philosophy of language as a distinct discipline has been modern linguistics, in particular generative grammar in the tradition of Chomsky that became the dominant paradigm in linguistics in the fifties and sixties of the previous century. And as it happens, both the generative tradition of Chomsky and analytic philosophy in its formal and less formal guises have been important factors in the development of formal semantics as well. Thus, it should come as no surprise that the two have something in common. That the communalities go beyond a common ancestry, but are reflected in substance and methods as well, will be argued later on.

However, be that as it may, it is still a good idea to keep philosophy of language and formal semantics separate, at least initially, since the role that is assigned to artificial languages and the ways in which these languages are employed in both does differ in a number of respects that are worth keeping in mind. Accordingly, we will begin with a

brief characterization of the way artificial languages play a role in philosophical investigations (section 2). Then we will look in some detail at their use in formal semantics, focusing on the philosophical presuppositions of the way in which they are employed there (section 3). The differences and resemblances are then the subject of section 4.

2 Philosophy

In philosophical research on language we mostly find artificial languages being used as notational aides, in much the same way as they are used in, for example, analytic metaphysics, in traditional epistemology, or in philosophy of science. The main goal of this use of artificial languages, such as those of first-order logic, modal logic, and the like, is to clarify a particular argument, or to precisely formulate a conjecture or a thesis. We could call this use of artificial languages that of "easing the conversation," as the purpose of it is to make certain elements in an argument or thesis more readily accessible and thereby more easy to evaluate critically (or more convincing as the case may be). Getting the relative scope of quantificational expressions ('everyone', 'most of us', 'some people',. . . 'always', 'sometimes', . . .) and modal operators ('necessarily', 'obliged', 'possibly', 'allowed' . . .), or the exact nesting of Boolean connectives and operators ('and', 'or', 'if . . . then . . .', . . ., 'not', . . .), right is often crucial for the understanding, and subsequent evaluation, of an argument, or for the grasp of the precise content of a particular thesis. Of course, in many cases the required disambiguation or explication can be provided without taking recourse to the use of formal notation, simply by explicitly stating, for example, the intended relative scope of two quantificational expressions. But such paraphrases may become somewhat cumbersome and complex, and anyway, the use of notation, apart from being helpful, also has a certain aesthetic appeal. A classical case where the use of notation helps to, if not solve, then at least make clear a particular problem is Aristotle's treatment of temporal modalities in the famous sea battle discussion in *Peri Hermeneias*, which has created a large number of diverging interpretations that in part revolve around the question what are the exact scopal properties that should be assigned to the relevant terms in Aristotle's text. Concretely, the question is whether Aristotle holds that the truth of a proposition about a future event depends on causal determinism (in which case we should construct Aristotle's premise as 'It is necessary that p or it is necessary that not-p'), or whether he merely subscribes in the relevant passages to the law of excluded middle (in which case the premise reads as 'It is necessary that (p or not-p)'.

Although this is certainly a proper use of artificial languages, we should also remark that it is not the most deep use. Be that as it may, what is more troubling is that in many cases authors that use a piece of notation forego the trouble of specifying what it means. And this actually may hamper understanding rather than facilitating getting a certain point across. In many cases authors rely on an implicit understanding on part of their readers of the notation they are using. Often, this may be excused: for instance, when one uses the quantifier-notation of first-order logic, it would be somewhat pedantic to explicitly state interpretation clauses for the quantifiers, given that there is an accepted, standard way of interpreting them. However, if one uses the box-operator from modal logic, for example, in a partial formalization of the premise or the conclusion of an argument, such a use really makes sense only if one also indicates which interpretation one assigns to it. With such a wide variety of modal systems around, one really can not rely on a common understanding on the part of all of one's readers.

This use of notational devices borrowed from various kinds of formal languages is quite akin to the way in which the use of mathematical notation has become part of our general conversational repertoire. We all employ Arabic numerals and simple arithmetical notation all the time, as if they were part of the natural language we use. This requires no knowledge or expertise beyond a basic level of practical ability. Certainly no acquaintance with anything as sophisticated as, for example, the Peano-axioms, or the Frege–Russell definitions of the natural numbers, is assumed here. This use of arithmetic is really all about easing the conversation, i.e., about making certain practical dealings with quantities and their basic properties more manageable.

Such use of notation borrowed from formal languages can very well be regarded as amounting to an extension of the natural language in which it is incorporated. Cf., the characterization that Wittgenstein gives in *Philosophical Investigations*, section 18 (Wittgenstein 2009):

> [. . .] ask yourself whether our language is complete;—whether it was so before the symbolism of chemistry and the notation of the infinitesimal calculus were incorporated in it; for these are, so to speak, suburbs of our language. (And how many houses or streets does it take before a town begins to be a town?) Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses.

For a more restricted set of users, logical notation, such as is borrowed from propositional or predicate logic, functions in much the same way. They use the notation to facilitate argumentation and debate, but not much hinges (in general, at least) on their having specific expertise in the logical systems that these languages are part of.

This is different in the context of the natural sciences, where the use of mathematics goes substantially beyond a 'mere' notational use. Here it is the formal systems as such, and not just the formal languages that provide the notational tools, that are being used. For example, the use of Riemann geometry in relativity theory is not a expedient notation but incorporates substantial claims about the nature of space and time. Hence, this formal system can be used to formulate explanations of phenomena and results that have been observed, and to deduce predictions that can be checked by further observation and experiment. The formal system here is more than a tool; it is part of the substance of the theory in question. Accordingly, here the use of a particular formal system is not just motivated by ease of use or similar concerns, it is subject to empirical validation or falsification, at least in principle. Inasmuch as the formal system is part of the theory, any experimental result or observed phenomenon that verifies or falsifies the theory reflects on the formal system by implication.

This seems to mark an important difference between the use of formal languages and systems in philosophy and in the sciences. That does not mean that in philosophy it is only the conversational and clarificatory use that we may encounter. Quite another use, less frequent but arguably more substantial, that is made of artificial languages in the context of philosophical discussions about language is when it is the properties of artificial languages themselves that are used as premises in some argumentation about natural language. Obviously, such a use is not of artificial languages as such (as is the case in the conversational use noted above) but of *systems,* i.e., of languages with an interpretation (which can be model-theoretic or proof-theoretic). An interesting example of such an

application is Putnam's appeal to the Löwenheim–Skolem theorem for first-order logic, as part of an argument against the possibility of providing natural languages with a referential semantics. (Cf., Putnam 1983; for a more concrete version of the argument, cf. Putnam 1981.) The theorem states that every countable first-order theory that has an infinite model has a model of size k for every infinite cardinal k. This means that such theories are not able to fix their models up to isomorphism. Putnam uses this result to argue that meaning does not fix reference, since given a set of true sentences in a given language we can always, by permutation of the objects in the domain, change the reference of subsentential terms without this affecting the truth values of the sentences in question. This shows, according to Putnam, that meaning as such does not fix reference.

It is important to note that an argument like this works on the basis of an assumption: that natural languages are sufficiently like formal languages in the relevant respects so as to allow a transfer of properties and results from formal languages (or systems) to natural languages. In this case, the assumption must be that any natural language 'contains' something like a first-order language and its associated model theoretic semantics. In many respects, this use of formal languages in philosophy of language is akin to the use that is made of, for example, epistemic logic, or probability theory, in formal epistemology. On the assumption that the phenomenon under consideration (natural language meaning, or epistemic reasoning) has a hidden structure that is sufficiently similar to what is explicit in some formal system (first-order logic, or epistemic logic), one projects results obtained with regard to the latter (such as (un)decidability) onto the former.

One of the problems in this type of use of formal languages is that the background assumptions (such as: 'natural languages contain first-order logic') are hardly ever discussed explicitly, let alone justified. But if we really use formal languages in this way, i.e., as formal systems that are part of a theory and that are instrumental in the deduction of certain conclusions, then we rely, not only on the applicability of the system as such but also on a prior decision to use this system rather than some other one. That priori decision of course calls for independent motivation. In the natural sciences such independent motivation comes from explanatory and/or predictive success with regard to independently acquired experimental and observational results. However, also in this type of philosophical application that is exactly what seems to be lacking.

Yet another use of artificial languages is when a particular language, or rather system, is appealed to as an arbitrator in order to decide certain issues that are in and of themselves not really linguistic in nature. For example, classical first-order logic is often held to encompass an appropriately parsimonious ontology, and thus is used as a standard to determine whether or not some entities that are assumed by some philosophical argument or theory actually are 'proper' entities. Quine's work on ontological commitment provides a classic example. This is a restrictive use, one that intends to single out a limited range of *bona fide* entities. On the other end of the spectrum we can locate positions that claim that whatever can be analyzed in some sufficiently general mathematical system, for example Zermelo–Fraenkel set theory (with or without the axiom of choice), or category theory, is something that one might justifiably appeal to. In both kinds of use, however different, it is the formal system that is used as a standard for what can, or cannot, be done in and through natural language. However, as in the previous kind of use, this type of argument really depends on the availability of external justification for the choice of the particular formal system that one employs as a standard. And, in such a case, such a justification seems to be lacking.

However, one might think it is formal semantics, as the empirical study of natural language meaning that uses similar formal languages—namely, those of logic and

mathematics—as its main tools, that actually provides such a justification. Therefore, it is now time to look at the use of artificial languages in that realm.

3 Semantics

The use of artificial languages in the study of the meaning of natural languages started taking off in a systematic manner only at the end of the 1960s. At the time, the rise of generative grammar in the Chomsky tradition had done much to expel philosophical doubts as to whether natural languages are systematic enough to allow the application of formal tools, such as those from model theoretic semantics, to them. Tarski had stated in his famous 1944 paper 'The Semantic Conception of Truth' that:

> The problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified. For other languages—thus, for all natural, 'spoken' languages—the meaning of the problem is more or less vague, and its solution can have only an approximate character. Roughly speaking, the approximation consists in replacing a natural language (or a portion of it in which we are interested) by one whose structure is exactly specified, and which diverges from the given language 'as little as possible.'

For some this meant that in order to become susceptible to any kind of formal treatment natural languages have to be reformed and regimented (beyond recognition, as some others complained). And yet other philosophers took the same 'observation' as evidence that any attempt to treat natural languages on a par with formal ones was wrong-headed to begin with, and that one would do better to study natural language in an informal, and much more descriptive manner. This approach, pioneered by Austin, Warnock, Ryle, and others, commonly known as 'ordinary language philosophy', was quite aptly also described as 'linguistic phenomenology.'

Both views, however, start from the assumption that natural languages are indeed not systematic enough to allow formal treatment, which is, of course, a complaint that has been leveled against natural languages by philosophers for centuries. The work of Chomsky in generative linguistics apparently inspired much more confidence in philosophers and logicians to assert that perhaps natural languages weren't as unsystematic and misleading as their philosophical predecessors had made them out to be after all.

To be sure, there had already been exceptions, such as Hans Reichenbach, whose *Elements of Symbolic Logic* (Reichenbach 1947) contains a large section devoted to the application of logic in the description of natural language phenomena, parts of which (in particular his treatment of the natural language tense system) became very influential much later on. And the potential relevance of Chomsky's work for philosophical semantics was already noted in 1953 by Yehoshua Bar-Hillel who wrote a programmatic paper in which he explores the possible connections. This attempt at cooperation met with a negative reaction from Chomsky himself, who throughout his entire career has remained hostile to the idea of some form of model theoretic semantics being of any relevance to linguistics. So, it is rather despite this reaction from the leader of the generative movement that at the end of the 1960s formal semantics began to flourish nonetheless.

A prominent example is Donald Davidson who claimed, in his seminal paper 'Truth and Meaning' from 1967, that:

> Philosophers of a logical bent have tended to start where the theory was and work out towards the complications of natural language. Contemporary linguists, with an aim that cannot easily be seen to be different, start with the ordinary and work toward a general theory. If either party is successful, there must be a meeting. Recent work by Chomsky and others is doing much to bring the complexities of natural languages within the scope of serious theory.

This sentiment is echoed in many of the early papers that constituted formal semantics as a discipline at the crossroads between linguistics, philosophy, and logic: generative linguistics shows how to capture the syntax of natural languages in a systematic and formal theory; in order to extend this to semantics, philosophy provides the necessary conceptual apparatus, consisting of analyses of meaning, reference, and truth that in many respects go back to the early days of analytic philosophy; and logic contributes the formal tools with which these concepts can be applied in a systematic fashion to natural language.

Of course, this cooperation between philosophers, linguists, and logicians took on different forms, and not everyone applied formal languages in formal semantics in exactly the same way. One of the most influential approaches turned out to be the one pioneered by Richard Montague. In his 1970 paper 'Universal Grammar', which outlines the theoretical machinery behind what became known as 'Montague grammar', he states:

> There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory. On this point I differ from a number of philosophers, but agree, I believe, with Chomsky and his associates.

This is a strong statement. No doubt Montague was well aware of the numerous differences that exist between the formal languages of logic and mathematics and natural languages, yet apparently he is also of the opinion that these are not 'important theoretical differences' that would constitute an obstacle to describing both with the same mathematical means.

Montague's statements suggest that formal languages and natural languages are on a par, but that is not how formal semanticists actually construe the relationship between the two. Their focus of interest is natural language meaning, and hence they use formal languages as tools in their investigations. The most explicit form this takes is when an interpreted formal language is used as a model for (some part of) natural language. The idea behind this methodology is that we can describe and explain the semantic properties of natural language expressions by setting up a systematic relationship between these expressions and those of some suitable formal language, and then use the semantics of the latter to act as a proxy for that of the former. This procedure of 'indirect interpretation', as it is often called, relies on the assumption that the formal system that is being used in this way somehow provides explanatory power with regard to the natural language. This approach was made popular by Montague's paper 'The Proper Treatment of Quantification in Ordinary English' (Montague 1973), but quite similar ideas can be found on early work of Lewis (1972), Cresswell (1973), and others.

This methodology has in essence remained unchanged from the early days of formal

semantics until pretty much the present-day. As then, formal semanticists model a particular phenomenon—for example, anaphora, or aspect, or vagueness—by either taking an existing formal system that has been developed independently and, in most cases, for different purposes, or by defining one themselves, and then use it to model the natural language phenomenon by providing a more or less strict and systematic translation of the relevant part of natural language into the language of the formal system. In the old days, the tendency was to stick to one particular formal system, that, of course, then had to have all the expressive power one could imagine one would need at some point. Montague used a system of intensional higher-order type theory, Cresswell went all the way and just used ZFC set theory; Davidson chose to restrict himself, for independent philosophical reasons, to the use of standard first-order logic. In particular in linguistic circles, Montague's choice initially won the day, but over the years subsequent developments have seen a plethora of formal systems being used by formal semanticists, unfortunately more often than not without much attention being paid to the overall compatibility of these systems. Domain theory, property theories, belief revision systems, event calculus, different many-valued logics, various nonmonotonic logics, dynamic logic, various forms of game theory, second-order type theory, Martin-Löf's type theory, untyped lambda-calculus, Boolean algebras, lattices of various kinds, set theory with or without ur-elements—basically everything in the book has been thrown at natural language phenomena at some point. And then there are the 'custom built' systems, such as various systems of discourse representation theory, or situation theory.

What is interesting about the use of this wide variety of formal systems is how the choices are motivated. In many cases, especially those where the semanticists use existing frameworks, they aren't motivated, at least not explicitly. In such cases one rather gets the impression that the formal system that is being employed isn't so much chosen for any particular properties it may have, but simply because 'it gets the job done'. For that to be the case it must of course have enough, and the right, expressive power, but in most cases this is, as it were, settled on incidental, almost ad hoc grounds, not on the basis of a prior investigation of the properties of the system in question. In some cases, though, the latter do play a role and may actually turn up in the motivation for the use of a particular system. Relevant considerations here usually center around properties such as (un)decidability, (in)completeness with respect to some proof system, expressive power (for example, first-order versus higher-order quantification), and similar concerns. These are, of course, properties of formal systems as such, not necessarily of the natural languages that are studied with the aid of them. However, the questions and concerns that are raised almost always relate to the natural languages, not to the formal systems. Thus, a decision to choose a formal system can be found to be motivated by concerns about the learnability of the natural language, or about its effective, practical applicability in, say, common sense reasoning. The assumption behind this is, then, that the relevant properties of the formal system that is used in the analysis of the formal language can, in some way, be translated back to the natural language, and in that way address the kind of concerns just mentioned.

Thus, artificial languages and formal systems are being used in the analysis of the semantics of natural languages in different ways and with different motivations. However, what appears to be a common assumption is that the formal system is regarded as a model of the natural language. In actual analyses and descriptions it is, of course, never considered to be a model of an entire natural language, since it is always only certain

aspects of a language (some class of expressions, a particular type of construction) that are under discussion. But in each case what is being analyzed and described is taken to be modeled by the formal system in the sense that all the relevant properties of the natural language are assumed to be adequately represented by properties of the formal system. For it is only on such an assumption that it makes sense to think of the formal system as a representation of the relevant semantic properties of the natural language.

This is a widely held view, most of the time simply assumed, sometimes explicitly stated. Textbooks are often the best source of statements of such basic points of view. The following is a quotation from a introduction to natural language semantics by Henriëtte de Swart (De Swart 1998):

> Given that direct interpretation of natural language with respect to the outside world (or some model of it) is not always easy, many semanticists opt for the indirect approach. We know that a translation can sometimes help us to determine the meaning of an expression. Suppose I speak French, but you don't, and we both speak English. In that case, I can teach you something about the meaning of a French expression by translating it into English [. . .] The same 'trick' can be used with the translation of natural language into some formal language. Suppose I can describe the meaning of an English expression by translating it into an expression of a formal language. Because there will be a full and explicit interpretation procedure for the expressions of this formal language, I will immediately have grasped the meaning of the English expression. Of course, I will only have access to it in an indirect way, namely via a translation procedure, but as long as the translation is perfect, the exact meaning will be captured.

This clearly illustrates that the idea of formal languages as models of natural languages is built right into the core methodology of formal semantics: namely, that of studying natural language by providing translations of relevant fragments into the language of some formal system. The same idea can be found in other textbooks (Cf., for example, Chierchia & McConnell-Ginet (2000):

> Is it possible to regard logical form [. . .] as providing us with a theory of semantic interpretation, with a theory that characterizes what we grasp in processing a sentence? [. . .] We think it is possible, as our logical forms do meet the main requirements that semantic representations are generally expected to meet.

However, it relies on two assumptions with regard to its object of study and the way in which this can be accessed. It is clear that using formal systems to study natural language meaning in this way, i.e., by devising translations between the two, works only if we can assume that both the meanings of the expressions of the formal language as well as those of the natural language are determinate and available prior to the analysis being carried out. They are determinate in the sense that they are able to guide the translation, in the sense of providing the necessary criteria for determining when the translation is actually correct. And they are available in the sense that as semanticists we have access to them independent of and prior to the use to which we put them. (For a more systematic analysis of the determinacy and availability assumptions, including their role in concerns with language reform in the work of Frege, Russell, early Wittgenstein, and others, see Stokhof 2007.)

Notice that we need these determinacy and availability assumptions with regard to both the meanings of the natural language expressions to be analyzed as well the meanings of the formal language that are used. Both are needed for otherwise there would be no way to judge the correctness of the translation, and this is essential if the methodology of indirect interpretation via translation is to succeed. This means that prior to the specification of the translation both are assumed to be given: we need to know what the meanings of the natural language expressions are and what meanings are assigned to the formal language expressions before we can start defining the translation or judge the correctness of any attempts. Consequently, what such an indirect specification of meaning does, at best, is to represent them in another way than the original natural language expressions themselves do. This may very well lead to a more perspicuous representation, or one that has other, technical advantages, but the one thing this methodology will not do is to actually provide meanings for natural language expressions, nor will it allow us to discover their formal properties, because prior to the analysis these are assumed to be determined and to be already available.

Where does this leave us with regard to how and why formal languages are used in semantics? For that we need to go back to philosophy.

4 Semantics and philosophy

The way in which interpreted formal languages are used in formal semantics to model relevant aspects of the semantics of natural language has much in common with the 'Putnam'-type applications in philosophy of language. Properties of the formal system are used to explicate and explain certain features of natural language meaning, which is made viable to begin with by the assumption of sufficient similarity between the two domains. As we noted above, this calls for independent motivation and justification, which philosophy appears not quite able to provide. In the sciences, this is different: there, data from experiments and observation can be used to test whether the assumed similarity between the formal system and the empirical domain actually obtains. This works, first of all, because the data used for testing can be obtained independently (at least in principle), and, secondly, because the formal system is integrated in the theory about the empirical domain.

The question now is whether in formal semantics independent justification of these assumptions can be procured as well. If so, then formal semantics is like the sciences and it may even provide the needed empirical justification for some of the philosophical analyses that make use of a similar methodology. If not, then formal semantics seems to be in the same boat as philosophy.

That the very methodology of formal semantics rests on the assumptions of determinacy and availability is ample reason to think that the required independent and prior justification is lacking here: the method of modeling natural language meaning by means of translation into expressions of an interpreted formal language assumes rather than justifies the existence of the required similarities between the former and the latter. So by itself this use of formal systems does not show that it is adequate, nor do the predictions and explanations that are based on it.

This point is reinforced by the observation that, unlike in the sciences, where the formal system that is being employed is built right into the heart of the theory, in formal semantics no such intrinsic link seems to exist. In view of the wide variety of formal systems being used, and taking into account that the choice of formal systems does not

appear to be limited by any substantial empirical argument, we must conclude that these systems play a very different role.

What might that role be? It seems we can make the following fundamental distinction between two ways of looking at what we do when we apply artificial languages in the study of natural language. First of all, we can look at a formal language as a model of a natural language. This is the traditional, and still dominant, view on the matter that we have outlined above, and which encounters the problems of justification that we also discussed. Second, we may consider a formal language as a tool—one of the many tools, we should add—in the study of natural language meaning. We can use a particular formal language to display certain inference patterns involving generalized quantifiers, say, and use another one to deal with, for example, the division between lexical and world knowledge. We can employ the compositionality with which we endow our formal languages to test whether a certain fragment of a natural language allows for a similarly compositional description. (But note that much here depends on what we have taken compositionality of natural languages to consist in to begin with, a question that arguably is not straightforwardly empirical.) What is crucially different on this second approach is that the adequacy criteria for our choice and employment of a particular formal language have to come from elsewhere: They derive from the practical concerns that we have, which by themselves are completely agnostic with respect to the tools that we may use or need to use. (And these concerns are even agnostic as to whether logical languages are the best tools. That, then, also becomes a matter that is up for discussion: For some purposes stochastic tools may be arguably better.)

On this second approach we employ formal languages in the study of natural languages, not because 'there is no important theoretical difference' between the two but because they are useful tools. A formal language is not a model of a natural one, but rather a tool that can be used to provide a 'perspicuous representation' of some part or aspect of it. In quite a similar vein one might look at the use of formal languages in philosophical analysis of language: There, too, the primary aim that is served by their employment is perspicuity and clarity. Of course, the content in both is different: Formal semantics deals with empirical phenomena, philosophical analysis with conceptual considerations. But the use both make of formal languages as a means for perspicuous representation is basically the same.

References

Chierchia, G. & McConnell-Ginet, S. (2000) *Meaning and Grammar*, Cambridge, MA: MIT Press

Cresswell, M. (1973) *Logics and Languages*, London: Methuen

Davidson, D. (1967) "Truth and Meaning," *Synthese* 17

Lewis, D. (1972) "General Semantics," *Synthese* 22:18–67

Montague, R. (1970) "Universal Grammar," *Theoria* 36: 373–98

Montague, R. (1973) "The Proper Treatment of Quantification in Ordinary English," in J. Hintikka, J. Moravcsik, and P. Suppes (eds.) *Approaches to Natural Language*, Dordrecht: Reidel

Putnam, H. (1981) "A Problem About Reference" in *Reason, Truth and History*, Cambridge: Cambridge University Press

Putnam, H. (1983) "Models and Reality," in *Realism and Reason*, Cambridge: Cambridge University Press

Reichenbach, H. (1947) *Elements of Symbolic Logic*, New York: Macmillan

Stokhof, M. (2007) "Hand or Hammer? On Formal and Natural Languages in Semantics," *The Journal of Indian Philosophy*, 35(5–6): 597–626

Swart, H. de (1998) *Introduction to Natural Language Semantics*. Stanford: CSLI

Wittgenstein, L. (2009) *Philosophische Untersuchungen / Philosophical Investigations*, Oxford: Blackwell, 4th revised edition

4.6
THE ROLE OF INTUITIONS

Michael Devitt

1 The Received View

How should we go about finding the truth about language? The received view is that we should proceed by consulting our intuitive judgments about language, our "intuitions." Indeed, it would be hard to exaggerate both the apparently dominant role of such intuitions in the philosophy of language and the agreement among philosophers that these intuitions should have this role. This emphasis on intuitions reflects, of course, a widely held view about the methodology of "armchair philosophy" in general.

Saul Kripke's *Naming and Necessity*, one of the most influential works in the philosophy of language, is often, and rightly, cited as an example of heavy reliance on intuitions. And Kripke is explicit that intuitions should have this important role:

> Of course, some philosophers think that something's having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking.
>
> (Kripke 1980: 42)

Stephen Neale gives another statement of the received view:

> Our intuitive judgments about what *A* meant, said, and implied, and judgments about whether what *A* said was true or false in specified situations constitute the primary data for a theory of interpretation, the data it is the theory's business to explain.
>
> (Neale 2004: 79)

These statements are quite general. Here are some typical claims about particular topics in the theory of language:

> *Semantic Content*: Herman Cappelan and Ernie Lepore attribute to a range of philosophers versions of the following, which they regard as the "Mistaken Assump-

tion": "A theory of semantic content is adequate just in case it accounts for all or most of the intuitions speakers have about speech act content, i.e., intuitions about what speakers say, assert, claim, and state by uttering sentences" (Cappelan and Lepore 2005: 53).

Reference: Stephen Stich thinks that the following view is "favored, albeit tacitly, by most philosophers": "the theory of reference is attempting to capture the details of a commonsense theory about the link between words and the world" where that theory involves, at least, a generalization of referential intuitions (Stich 1996: 6).

Truth Conditions: Jason Stanley and Zoltan Szabó endorse the view that "accounting for our ordinary judgements about the truth-conditions of various sentences is the central aim of semantics" (Stanley and Szabó 2000: 240).

What Is Said: Robyn Carston thinks that the various criteria in the pragmatics literature for placing "pragmatic meanings" into "what is said", "in the end, . . . all rest . . . on speaker/hearer intuitions" (Carston 2004: 10). François Recanati claims that "'what is said' must be analysed in conformity to the intuitions shared by those who fully understand the utterance" (Recanati 2004: 14).

Syntax: Noam Chomsky claims that 'linguistics . . . is characterized by attention to certain kinds of evidence . . . largely, the judgments of native speakers' (Chomsky 1986: 36). Liliane Haegeman, in a popular textbook, says that 'all the linguist has to go by . . . is the native speaker's intuitions' (Haegeman 1994: 8).

2 The Task?

It is clear from these claims, and other similar ones, that intuitive judgments are commonly thought to provide *the evidence* or, at least, *the main evidence* for theories of language. Yet, often, such claims seem to suggest a stronger view: *the very task* of theories of language is to explain or systematize competent speakers' intuitions about language.

This view of the task is very puzzling. For, the obvious way to describe the task of the theory of language is to *explain the nature of language*, to explain properties like *meaning*, *truth*, *reference*, and *grammaticality*—real properties of linguistic expressions playing some sort of explanatory role. If we start from this view, surely as good a starting place as one could have, why take the task to be to capture *ordinary intuitions about* such properties, intuitions that must simply reflect folk theory? That would seem to be appropriate only if we assume that the folk must be right about language. But why assume that? We don't suppose that the folk are authorities on physics, biology, or economics; why suppose that they are authorities on linguistics and semantics?

So let us set aside the idea that the study of linguistic intuitions is the task of theories of language and turn to the more plausible view that these intuitions are evidence for those theories.

3 Experimental Semantics

The evidential role of intuitions in semantics, indeed, the evidential role of intuitions in philosophy generally, has recently grabbed the attention of a group known as "experimental philosophers." They have noted that the intuitions that play this evidential role are usually not those of the folk but rather those of the philosophers themselves. In response, they have conducted experiments to test whether the folk share these intuitions. Thus, in a well-known experiment, the intuitions that Kripke airs about reference were tested against those of undergraduates in Rutgers and Hong Kong (Machery et al. 2004). The results raised serious doubts in the experimenters' minds about Kripke's refutation of the description theory of names. There is plenty of room for debate about whether, and to what extent, the results *should* raise these doubts (Martí 2009, Devitt 2011a, Ichikawa et al. 2011). But one thing such experiments surely should do is focus our attention on two important questions: Why is the evidential role given to intuitions thought to be appropriate? Is it really appropriate? These questions will be the focus of this chapter.

4 "Cartesianism"

It would, of course, be appropriate to give a person's intuitions an important evidential role if we could be confident that they reflected *knowledge*. And the received view is that a competent speaker of a language does indeed have knowledge about her language, *propositional* knowledge, "tacitly" at least, *simply in virtue of being competent* in the language: "It is an undeniable feature of the notion of meaning . . . that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same" (Dummett 1978: 131). "The natural view is that one has *some kind of* privileged semantic self-knowledge" (Loar 1987: 97). The idea of this sort of privileged access—that we are in a special position to know about our own competence—is an instance of general "Cartesianism":

> Since Descartes, it has seemed undeniable to most philosophers that each of us has a privileged way of knowing about his or her own mental states. . . . whenever we have a thought, belief, intention, or desire, we can in principle come to know *what* we think, believe, intend, or desire just by internal examination, without engaging in an empirical investigation of the external world.
>
> (McKinsey 1994: 308)

The idea that we have a Cartesian access to semantic facts seems to be an almost unquestioned part of the semantic traditions of Frege and Russell. Consider, for example, the assumption that our competence consists in propositional knowledge of truth conditions. Herbert Heidelberger (1980) has shown how widespread this assumption is with references to Wiggins, Strawson, Davidson, Frege, Wittgenstein, Quine, and Carnap. He points out that it seems to be regarded as "uncontroversial . . . harmless . . . perhaps unworthy of serious discussion" (p. 402). Gareth Evans says, "perhaps no one will deny it" (1982: 106).

It is worth noting that one of the most famous arguments in the philosophy of language seems to rest on Cartesiansim. This is the Fregean argument that 'Hesperus = Hesperus' and 'Hesperus = Phosphorus' must differ in meaning, hence 'Hesperus' and

'Phosphorus' must differ in meaning, because they differ in cognitive significance or informativeness. But why would the latter *epistemic* difference be taken to establish the former *semantic* difference? The reason, I suggest (Devitt 1996: 172–3), is that it is taken for granted that two sentences mean the same only if competent speakers *know* that they do; see Dummett above. And speakers could know this only if sentences did not differ in cognitive significance for them. For, if they differed in significance, speakers would not think that they were synonymous.

5 A Priori Knowledge?

Why should we suppose that ordinary competent speakers have this knowledge of semantic facts? Many seem to think that the knowledge is a priori. Thus Jerrold Katz claims: "We know sense properties and relations of expressions on the basis [of] the speaker's a priori linguistic intuitions in clear cases" (Katz 1997: 21). And Michael McKinsey thinks that it is "fairly clear" that "the principle that the meanings of words are knowable a priori . . . is taken for granted by most philosophers of language and by many linguists" (McKinsey1987: 1).

Now, of course, the idea that some knowledge is a priori is widespread in philosophy. Nonetheless, Quine has raised serious doubts about it. The main problem with the idea, in my view (Devitt 1994, 1996; 1998; 2010a), is that we do not have even the beginnings of an account of what a priori knowledge *is*. We are simply told what it *isn't*: namely, empirical knowledge. Still, suppose we set such general doubts aside and accept that at least our knowledge of mathematics and logic is a priori, what could be the basis for supposing that our knowledge of *meanings* is too? The meaning of a word is presumably constituted by relational properties of some sort: "internal" ones involving inferential relations among words and/or "external" ones involving certain direct causal relations to the world referred to. Where the meaning is partly constituted by a certain external relation—plausible examples are the meanings of 'Cicero', 'elm', and 'water'—then it is hard to see how a priori reflection on what is "inside the head" could establish that such a relation constituted a meaning. But even the meaning-constituting internal relations pose serious problems for the a priori view. Let Susan be a competent speaker alleged to have a priori knowledge that, say, the inferential relation between 'bachelor' and 'unmarried' is part of the meaning of 'bachelor'. The first problem for the a priori view is that it requires that Susan comes to this belief about meaning rather than the belief that the inferential relation is simply the reflection of familiar knowledge about bachelors. Why should we suppose that her competence alone leads her to make this apparently large *theoretical* step? Suppose, however, that she does make it. The second and more serious problem for the a priori view is that we seem to have no basis for thinking that, simply in virtue of her competence, Susan's belief about meaning is *justified*. We have no basis for giving her belief any special epistemic authority, and thus turning it into *knowledge*. We need a plausible explanation of these allegedly nonempirical processes of belief formation and justification and some reasons for believing in them.

6 Embodied Theory?

If the view that competent speakers have a priori knowledge of linguistic facts does not hold up, what else could justify the ubiquitous reliance on intuitions in the philosophy of language? Philosophers provide little in the way of an answer. As Jaakko

Hintikka remarks, talking of philosophy generally: "One searches the literature in vain for a serious attempt to provide" a justification for the appeal to intuitions (Hintikka 1999: 130). In similar vein, Timothy Williamson remarks: "there is no agreed or even popular account of how intuition works, no accepted explanation of the hoped-for correlation between our having an intuition that P and its being true that P." He describes this as "a methodological scandal" (Williamson 2007: 215). Still, we can look to linguistics for a possible answer for *semantic* intuitions. (Perhaps we should see this answer as an explication of the obscure a priori answer.)

The common linguistic view of intuitive judgments is expressed in passages like the following: "It seems reasonably clear, both in principle and in many specific cases, how unconscious knowledge issues in conscious knowledge . . . it follows by computations similar to straight deduction" (Chomsky 1986: 270)

I have described the common view as follows: linguistic competence, all on its own, "provides information about the linguistic facts. . . . So these judgments are not arrived at by the sort of empirical investigation that judgments about the world usually require. Rather, a speaker has a privileged access to facts about the language, facts captured by the intuitions, simply in virtue of being competent . . ." (Devitt 2006a: 483–4; 2006b: 96)

On this view, intuitive syntactic judgments are, "noise" aside, "the voice of competence." Let's call this thesis "VoC." We can identify two versions of it, a "standard" one which requires linguistic rules (and principles) to be *represented* in the mind and a "nonstandard" one which does not. According to the standard version, suggested by the Chomsky quote, speakers derive their intuitive judgments from their representations of linguistic rules by a causal and rational process like a deduction. According to the nonstandard version, the intuitions must be provided somehow by embodied but unrepresented rules (Devitt 2006a: 482–6; 2006b: 96–8).

Stich has suggested that philosophers of language might be guided by linguistics in seeking a justification for the authoritative role given to semantic intuitions (Stich 1996: 40; see also Hintikka 1999, Williamson 2007). Stich's particular concern is with referential intuitions. He suggests that philosophers may think that speakers derive their referential intuitions from a representation of referential principles. So, just as, according to the standard version of VoC, the true grammar that linguists seek to discover is already represented in the mind of every speaker, so too, according to this suggestion, are true semantic theories of reference, meaning, and the like. Semantic intuitions, like syntactic ones, are the result of something like a deduction from a represented theory. Thus, speakers have Cartesian access to linguistic facts simply in virtue of being competent.

Stich's suggestion concerns the standard version of VoC, which posits representations of theories of language in the minds of ordinary speakers. What about the nonstandard version which does not posit these representations? Stich does not consider this version as a model for semantic intuitions, but if the model worked for syntax it should work for semantics as well.

Someone who took the usual linguistic view of grammatical intuitions might well be tempted by this analogous view of semantic intuitions. So, it is interesting to note that Chomsky is not tempted. He expresses skepticism about "contemporary philosophy of language" and its practice of "exploring intuitions about the technical notions 'denote', 'refer', 'true of', etc." He claims that "there can be no intuitions about these notions, just as there can be none about 'angular velocity' or 'protein'. These are technical terms

of philosophical discourse with a stipulated sense that has no counterpart in ordinary language." (Chomsky 1995: 24)

So Chomsky is skeptical about the use philosophers make of semantic intuitions. But he is not, of course, similarly skeptical about the use linguists make of syntactic ones. Why the difference? If skepticism about semantic intuitions is appropriate, then surely just the same skepticism is appropriate about the syntactic ones, and for just the same reason. All the terms in linguistic theory are, in the relevant sense, technical and theory-laden. A few like 'grammatical' and 'sentence' have counterparts in ordinary language but so too do 'denote' and 'refer'. Semantic and syntactic intuitions are on a par. Chomsky seems to have given a good reason for rejecting VoC altogether.

In the last section we looked dimly on the idea that linguistic intuitions are a priori. We have just raised a preliminary doubt about VoC. Before raising more doubts, we need to describe an alternative view. This alternative accompanies the view that linguistic competence is simply a skill.

7 Competence as a Skill

It is natural to follow the folk in talking of a person's linguistic competence as a form of "knowledge." We noted in section 4 that it is common to think of this knowledge as *propositional*, knowledge-*that*, leading to the idea that speakers have Cartesian access to linguistic facts. But there is an alternative. If we must follow the folk in talking of competence as knowledge, we can think of it as mere knowledge-*how*. However, I think it is better to talk of the competence as simply a skill or ability.

What skill or ability is the competence? Accepting, as we should, that "language expresses thought," I give the following answer:

> the competence is the ability to use a sound of the language to express a thought with the meaning that the sound has in the language in the context of utterance; and the ability (together with some pragmatic abilities) to assign to a sound a thought with the meaning that the sound has in the language in the context of utterance (similarly for inscriptions, etc.).
>
> (Devitt 2006b: 148)

We can move to a more theory-laden view of competence if we adopt the popular, and in my view correct, representational theory of the mind according to which any thought involves standing in a certain functional relation to a mental representation. Competence is then "the ability to *translate* back and forth between meaningful mental representations and the sounds of the language." And if we go further to the controversial language-of-thought hypothesis according to which the mental representation is language-like, the translation is "between mental *sentences* and the sounds of the language" (ibid.). Finally, linguistic competence is complex, consisting of syntactic competence and lexical competence. Thus, going along with the language-of-thought hypothesis, syntactic competence is the ability to translate back and forth between the syntactic structures of the sounds of the language and the structures of mental sentences. And lexical competence is the ability to translate back and forth between the words of the language and mental words.

Why think that that linguistic competence is just a skill or ability? Briefly, because it has all the marks of one: It has limited plasticity; it is extraordinarily fast; the process

of exercising it is unavailable to consciousness; once established, it is "automatic" with the result that it can be performed while attention is elsewhere (Devitt 2006b: 209–10); it is very likely acquired by "implicit learning" (Devitt 2006b: 219). But shouldn't we suppose that in the case of linguistic competence, the skill involves knowledge-that? I argue not (Devitt 2006b). We should suppose that only if we have some powerful reasons for doing so. Otherwise the supposition seems gratuitous. Why suppose that simply in virtue of being competent in a language a person must have propositional knowledge about the language? Why suppose that speakers have this sort of Cartesian access to linguistic facts? Why not suppose, rather, the modest view that any knowledge of these facts that a speaker may have comes from ordinary empirical reflection on linguistic phenomena?

Assumptions that competence involves propositional knowledge are very immodest. One would expect, therefore, that they would be well supported by arguments. This is not what we find. Arguments for them are few and far between and remarkably thin. (Stanley and Williamson 2001 may rate as an exception; Devitt 2011b is a response.) One gets the impression that these propositional assumptions are thought to be too obvious to need argument.

8 Intuitions As Empirical Judgments

If linguistic competence is simply a skill, requiring no propositional linguistic knowledge, it is hard to see how it could be the source of our linguistic intuitions as VoC requires. But what is the alternative? I think that we should see intuitions as ordinary empirical judgments (Devitt 1994; 1996; 2006a,b,c; 2011a; see also Kornblith 1998). Like the skill-view of competence, this alternative is modest: It makes do with cognitive states and processes we were already committed to.

The alternative is based on a view of intuitions *in general*: They "are empirical theory-laden central-processor responses to phenomena, differing from many other such responses only in being fairly immediate and unreflective, based on little if any conscious reasoning" (Devitt 2006a: 491; 2006b: 103). From this perspective, we should trust a person's intuitions to the degree that we have confidence in her empirically based expertise about the kinds under investigation. Sometimes the folk may be as expert as anyone: Intuitions laden with "folk theory" are the best we have to go on. Perhaps this is the case for a range of psychological kinds. For most kinds, it clearly is not: We should trust intuitions laden with established scientific theories. Consider, for example, a paleontologist in the field searching for fossils. She sees a bit of white stone sticking through gray rock, and responds immediately, "a pig's jawbone." We trust her intuitive judgment in a way that we would not trust folk judgments because we know that it is the result of years of study and experience of old bones; she has become a *reliable indicator* of the properties of fossils. Similarly, we trust the intuitions of the physicist over those of the folk about many aspects of the physical world where the folk have proved notoriously unreliable. And recent experiments have shown that we should have a similar attitude to many psychological intuitions. Thus, the cognitive psychologist, Edward Wisniewski, points out that:

> researchers who study behavior and thought within an experimental framework develop *better* intuitions about these phenomena than those of the intuition researchers or lay people who do not study these phenomena within such

> a framework. The intuitions are better in the sense that they are more likely to be correct when subjected to experimental testing.
>
> (Wisniewski 1998: 45)

Although we may often be right to trust an intuition in the short run, it is crucial to see that nothing rests on it in the long run. We can look for more direct evidence in scientific tests. In such a scientific test we examine the reality the intuition is *about*. These scientific examinations of reality, not intuitions about reality, are the primary source of evidence. The examinations may lead us to revise some of our initial intuitions. They will surely show us that the intuitions are far from a complete account of the relevant bit of reality.

What then should we make of *linguistic* intuitions? All these intuitions arise from reflection on linguistic data (understood as pieces of linguistic usage, on the model of "primary linguistic data"). The competent speaker has ready access to a great deal of linguistic data just as the competent typist has to a great deal of typing data and the competent thinker has to a great deal of thinking data: The competent speaker and her competent fellows produce linguistic data day in and day out. So she is surrounded by tokens that may, *as a matter of fact*, refer to so and so, be true in such and such circumstances, be grammatical, be ambiguous, and so on. So she is in a position to have well-based opinions about language by reflecting on these tokens. This is not to say that she will reflect. Indeed, a totally uneducated person may reflect very little and hence have few if any intuitive judgments about her language. Still it is clear that the normal competent speaker with even a little education *does* reflect on linguistic reality just as she reflects on many other striking aspects of the world she lives in. And this education will usually provide her with the terms and concepts of folk semantics and linguistics, at least. As a result she is likely to be able to judge in a fairly immediate and unreflective way that a token *does* refer to so and so, *is* true in such and such circumstances, *is* grammatical, *is* ambiguous, and so on. Such intuitive opinions are empirical central-processor responses to linguistic phenomena. They have no special authority: although the speaker's competence gives her ready access to data it does not give her Cartesian access to the truth about the data.

Still, are these intuitions likely to be right? I think we need to be cautious in accepting them: Thinking about language is notoriously hard and the folk are a long way from being experts. Still it does seem to me that their "simplest" intuitions, involving syntactic and semantic vocabulary that we suppose the folk have mastered well enough, are quite likely to be right. So we can often be confident about judgments that a name "refers" to *x*, that this pronoun must "refer to the same thing" as that name, and that this expression is "ambiguous." In sum, we have good reason to suppose that the core judgments of folk linguistics, reflecting the "linguistic wisdom of the ages," are good evidence for linguistic theories.

This having been said, the intuitions that philosophers and linguists should prefer are the ones that they do, as a matter of fact, mostly prefer: those of philosophers and linguists themselves. For, they are much more expert. This is particularly so when we get beyond the simple cases to fanciful ones like Kripke's Gödel (1980) and Putnam's Twin Earth (1975). Just as the intuitions of paleontologists, physicists, and psychologists in their respective domains are likely to be better than those of the folk, so too the intuitions of philosophers and linguists (c.f. Machery et al. 2004).

To say that intuitions, whether those of philosophers or of the folk, are evidence is not to say that they are the only, or even the primary, evidence. Indeed, we can look for

more direct, less theory-laden, evidence by studying what the intuitions are *about*, the linguistic reality itself. But how do we do that? I shall address this briefly in section 10.

9 Rejecting VoC

I think (Devitt 2006a,b,c) that there are several reasons for preferring this modest explanation of intuitions to VoC, whether as a proposal for syntactic intuitions or as one, following Stich's suggestion, for semantic intuitions. (For discussion, see Textor 2009, Fitzgerald 2010, Devitt 2010c, d.) I shall mention the main ones.

The main consideration against the *standard* version of VoC is as follows. It is of course *possible* that the competent speaker's intuitions about her language are reliable because they are derived from her mental representations of theories of the language. It has been argued, mistakenly in my view, that we need to posit such representations to explain language acquisition and use. In any case, we surely do not need the posit to explain the reliability of intuitions about language. Consider the analogous phenomena for typing and thinking. We can explain the reliability of intuitions about those processes by adverting to cognitive states and processes that we are already committed to, without positing representations of the rules that govern the processes. These modest explanations seemed perfectly adequate for the job and, indeed, much more plausible than their representational rivals. So do the similarly modest explanations in the linguistics case. Language is an important part of the human environment. It is not surprising that empirical reflection on linguistic data, aided by some education, should make people fairly reliable detectors of the most obvious facts about language. We are surely similarly reliable about other important parts of the environment: for example, the physical, biological, and psychological parts.

The main consideration against the *nonstandard* version of VoC is that we do not have *any idea* how embodied but unrepresented rules might provide linguistic intuitions. Not only do we lack the details needed for a plausible explanation, but attention to other similar systems gives good reason to suppose that the linguistic system does not provide these intuitions and so we *could never* have the details. The explanation would require a relatively direct cognitive path from the embodied rules of the language to beliefs about expressions of that language, a path that does not go via central-processor reflection on the data. What could that path be? Consider some other examples. It is very likely that rules that are embodied but not represented govern our swimming, bicycle riding, catching, typing, and thinking. Yet there does not seem to be any direct path from these rules to relevant beliefs. Why suppose that there is such a path for linguistic beliefs? Why suppose that we can have privileged access to linguistic facts when we cannot to facts about these other activities? We do not have the beginnings of a positive answer to these questions and it seems unlikely that the future will bring answers.

10 Other Evidence

On the view I have presented linguistic intuitions are fallible empirical judgments that are, at best, only indirect evidence of linguistic reality. Where can we find more direct evidence? This is a large question, beyond the scope of this chapter. I shall briefly mention some points made elsewhere (Devitt 1994, 1996, 2006a, b, 2008, 2011a; see also Martí 2008).

We should start by identifying the "linguistic reality" in question. This reality is to be found in the utterances of language users. But what is it about these utterances that is theoretically interesting? It is natural to think that our interest is in their *meanings*. Related to this, since "language expresses thought," it is natural to think that we are interested in the meanings (contents) of thoughts. But this talk of meanings is sadly vague, as many have noted: It is far from clear what counts as a meaning that needs explaining. We need to be much more precise in identifying the subject matter of semantics. We should identify the meanings of utterances and thoughts with certain properties of them that are crucial to their causal roles. Of particular interest here are the roles of thoughts and utterances in causing intentional behaviors and informing us about the world. So we can identify the reality that concerns us with the properties of thoughts and utterances in virtue of which they play those causal roles.

This reality can provide lots of direct evidence of its nature. The main source is "the corpus," the utterances that the folk have produced and are producing as they go about their lives. We can look at the circumstances and consequences of these productions. Indeed, the linguistic intuitions that are indirect evidence—for example, that people ignorant about Einstein nonetheless refer to him by 'Einstein' (Kripke 1980)—come largely from reflecting on these features of linguistic usage. The technique of "elicited production" could provide further direct evidence: Experimental situations "are designed to be uniquely felicitous for production of the target structure" (Thornton 1995: 140). Next, rather than *creating* situations in which we see what people say or understand, we can *describe* such situations and ask people what they *would* say or understand. Finally, and relatedly, we can look to ordinary opaque attitude ascriptions. Day in and day out, folk, and social scientists, use "content clauses," usually 'that' clauses like that in "Frank said that Madonna is gorgeous," to ascribe properties to thoughts and utterances for the very purpose of explaining behavior and learning about the world. Insofar as these ascriptions are right—and we have good reason to suppose that they mostly are right because they are mostly *successful*—they provide evidence of the nature of the reality that concerns us.

This direct evidence often involves behaviors that are aptly called "intuitive," but it is very different from the indirect evidence provided by the intuitions that that have concerned us here: For the behaviors are linguistic *performances*, whereas those intuitions are *judgments about* language.

11 Conclusion

Intuitive metalinguistic judgments seem to be the main source of evidence in the philosophy of language. We should reject the idea that this is justified because those intuitions are a priori or "the voice of competence." Rather, we should see the intuitions as empirical judgments. As such, they are often good, albeit indirect, evidence. And the intuitions we should prefer are those of philosophers and linguists, because they are more expert. However, we can find more direct evidence by looking to the linguistic reality that these intuitions are about.

Acknowledgements

Thanks to Fiona Cowie and David Sosa for helpful comments on a draft and to Elmar Geir Unnsteinsson for help with research.

Related Topics

4.1 The Role of Experiment
4.2 The Role of Linguistics
6.5 A Priority.

References

Cappelen, H., and E. Lepore (2005) *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*, Oxford: Blackwell Publishing.

Carston, R. (2004) "Truth-Conditional Content and Conversational Implicature," in C. Bianchi (ed.) *The Semantic/Pragmatics Distinction*, Standford: CSLI Publications.

Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origin, and Use*, New York: Praeger Publishers.

—— (1995) "Language and Nature," *Mind* 104: 1–61.

DePaul, M. R., and W. Ramsey (eds.) (1998) *Rethinking Intuition: the Psychology of Intuition and its Role in Philosophical Inquiry*, London: Rowan and Littlefield Publishers.

Devitt, M. (1994) "The Methodology of Naturalistic Semantics," *Journal of Philosophy* 91: 545–72.

—— (1996) *Coming to Our Senses:A Naturalistic Program for Semantic Localism*, Cambridge: Cambridge University Press.

—— (1998) "Naturalism and the A Priori," *Philosophical Studies* 92: 45–65. Reprinted in Devitt 2010b.

—— (2006a) "Intuitions in Linguistics," *British Journal for the Philosophy of Science* 57: 481–513.

—— (2006b) *Ignorance of Language*, Oxford: Clarendon Press.

—— (2006c) "Intuitions," in V. G. Pin, J. I. Galparaso, and G. Arrizabalaga (eds.) *Ontology Studies Cuadernos de Ontologia: Proceedings of VI International Ontology Congress (San Sebastian, 2004)*, San Sebastian: Universidad del Pais Vasco. Reprinted in Devitt 2010b.

—— (2008) "Explanation and Reality in Linguistics," *Croatian Journal of Philosophy* 8: 203–31

—— (2010a) "No Place for the A Priori," in M. J. Shaffer and M. Veber (eds.) *What Place for the A Priori*, Rodopi Press, forthcoming. Reprinted in Devitt 2010b.

—— (2010b) *Putting Metaphysics First: Essays on Metaphysics and Epistemology*, Oxford: Oxford University Press.

—— (2010c) "What 'Intuitions' are Linguistic Evidence?" *Erkenntnis* 73: 251-64.

—— (2010d) "Linguistic Intuitions Revisited," *British Journal for the Philosophy of Science* 61: 833-65.

—— (2011a) "Experimental Semantics," *Philosophy and Phenomenological Research* LXXXII: 418-35.

—— (2011b) "Methodology and the Nature of Knowing How," *Journal of Philosophy* CVIII: 205-18.

Dummett, M. (1978) *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press.

Evans, G. (1982) *The Varieties of Reference*, J. McDowell (ed.), Oxford: Clarendon Press.

Fitzgerald, G. (2010) "Linguistic Intuitions," *British Journal for the Philosophy of Science* 61: 123–60.

Haegeman, L. (1994) *Introduction to Government and Binding Theory*, 2nd edn, Oxford: Blackwell Publishers.

Heidelberger, H. (1980) "Understanding and Truth Conditions," in P. A. French, T. E. Uehling, Jr., and H. K. Wettstein (eds.) *Midwest Studies in Philosophy Vol. 5, Studies in Epistemology*, Minneapolis: University of Minnesota Press.

Hintikka, J. (1999) "The Emperor's New Intuitions," *Journal of Philosophy* 96: 127–47.

Ichikawa, J., I. Maitra, and B. Weatherson, B. (2011) "In Defense of a Kripkean Dogma," *Philosophy and Phenomenological Research*, forthcoming.

Katz, J. (1997) "Analyticity, Necessity, and the Epistemology of Semantics," *Philosophy and Phenomenological Research* 57: 1–28.

Kornblith, H. (1998) "The Role of Intuition in Philosophical Inquiry: An Account with no Unnatural Ingredients," in DePaul and Ramsey 1998.

Kripke, S. A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.

Loar, B. 1987 "Subjective Intentionality," *Philosophical Topics* 15: 89–124.

McKinsey, M. (1987) "Apriorism in the Philosophy of Language," *Philosophical Studies* 52: 1–32.

—— (1994) "Individuating Belief," in J. Tomberlin (ed.) *Philosophical Perspectives, 8: Logic and Language*, Atascadero, CA: Ridgeview Publishing Company.

Machery, E., R. Mallon, S. Nichols, and S. P. Stich (2004) "Semantics, Cross-Cultural Style," *Cognition* 92: B1–B12.

Martí, G. (2009) "Against Semantic Multi-Culturalism," *Analysis* 69: 42–48.

Neale, S. (2004) "This, That, and the Other," in M. Reimer and A. Bezuidenhout (eds.) *Descriptions and Beyond*, Oxford: Clarendon Press.
Putnam, H. (1975) "The Meaning of 'Meaning,'" in *Mind, Language and Reality: Philosophical Papers Vol. 2*. Cambridge: Cambridge University Press.
Recanati, F. (2004) *Literal Meaning*, Cambridge: Cambridge University Press.
Stanley, J., and Z. G. Szabó (2000) "On Quantifier Domain Restriction," *Mind and Language* 15: 219–61.
Stanley, J. and T. Williamson (2001) "Knowing How." *Journal of Philosophy* 98: 411–44.
Stich, S. (1996) *Deconstructing the Mind*, New York: Oxford University Press.
Textor, M. (2009) "Devitt on the Epistemic Authority of Linguistic Intuitions," *Erkenntnis* 71: 395–405.
Thornton, R. (1995) "Referentiality and Wh-Movement in Child English: Juvenile D-Linkuency," *Language Acquisition* 4: 139–75.
Williamson, T. (2007) *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.
Wisniewski, E. J. (1998) "The Psychology of Intuitions," in DePaul and Ramsey, 1998.

Section V

LOGIC FOR PHILOSOPHERS OF LANGUAGE

5.1
MODEL THEORY: WHAT IT IS AND WHAT IT ISN'T

John P. Burgess

Introduction

Model theory is the branch of logic concerned with relations between formal languages and extralinguistic structures. The most basic technical notion is that of a formula being *true* in a given structure, and the subject originated with Tarski and his famous definition of truth for classical first-order languages. On a common usage followed here, a *model* is just a structure of the kind sentences are true or false in. Model theory contrasts with proof theory, which is concerned only with formal deducibility relations among sentences, regardless of any connection with extralinguistic structures. Important post-Tarski contributors to the model theory of classical first-order languages have included Abraham Robinson, who gave the subject an orientation towards applications to abstract algebra, where the pertinent structures are groups and rings and fields, and Saharon Shelah, who introduced methods of great technical sophistication. Many mathematical logicians today use "model theory" to refer only to technically sophisticated, algebra-oriented, first-order model theory, a subject with little bearing on philosophy. Model theory has been developed also, however, for nonclassical formal languages (tense, modal, and other), especially by Saul Kripke and his successors. This branch of the subject claims applications to philosophy of language and theoretical computer science.

In the usage of computer science and philosophical logic, model-theoretic concepts are often called "semantic," and proof-theoretic concepts "syntactic." This usage can be traced back to Tarski, who noted that the word "true" was used in several senses, and to distinguish the one that concerned him drew on vocabulary used in Morris (1938), where *syntax* is concerned with relations among words, *semantics* with relations between words and things, and *pragmatics* with relations among words and things and people. Tarski was concerned with the notion of truth according to which "Snow is white" is true iff snow is white. William James, by contrast, took "Snow is white" to be true iff it is useful to believe that snow is white. (Here "iff" abbreviates "if and only if.") Tarski's notion is "semantic" in involving words and a thing (snow), while James's is "pragmatic" in involving also people (believers). Tarski called his development of a rigorous definition of truth as he conceived it "the establishment of scientific semantics."

Linguists, by contrast, apply the term "semantics" to the branch of their subject concerned with meaning in natural language. That there is a tension between this usage and the Morris–Tarski usage is evident if one considers the contrast between (i) denotation or reference or extension and (ii) connotation or sense or intension. Morris's words-and-things conception of semantics is more suggestive of (i), while the linguists' meaning conception of semantics is more suggestive of (ii).

It is hard to keep the two notions of "semantics" apart in theoretical computer science, since there one is concerned with artificial languages, and the specification of the conditions under which a sentence of such a language is to count as true in a given structure is about all one gets by way of a specification of what such a sentence is supposed to mean. Elsewhere it is crucial to distinguish, since fallacies of equivocation between the two senses of "semantics" can be damaging to the understanding of scope, limits, and status both of model theory and of meaning theory. For instance, confusion of "formal semantics" or model theory, to which Tarski's definition of truth is central, with "linguistic semantics" or meaning theory can lead straight to *Davidsonianism*, the contentious claim that specification of truth conditions is central to giving a theory of meaning for natural language; but to arrive at Davidsonianism *that* way would be to prejudge a controversy that can only be fairly judged after an examination of rival research programs and empirical linguistic data.

In logic, the nature of any relationship between models and meaning varies with the formalism under consideration (classical first-order, tense, modal, or other) and must be considered separately in each case. The methods of model theory are too flexible for it to be safe to assume that, just because there is a rigorously defined notion of model for a given formalism, it must have some intuitively intelligible meaning. And they are often too contrived for it to be safe to assume that, just because some feature is present in the models used for a given formal language, an analogous feature must be present also in whatever that formal language may be intended to represent. What follows here is an elaboration and illustration of such warnings.

Classical first-order logic

On one traditional conception, logic's central concern is with *consequence* or *implication*, where premises have as a consequence or imply a conclusion iff form alone guarantees that if the premises are all true, then the conclusion is true. The notion of *validity* is the special, zero-premise case: A conclusion is valid if its form alone guarantees that it is true. The general case can be reduced to the special case, since $A_1, \ldots, A_n$ imply B iff the conditional, sometimes called the "leading principle" of the inference from $A_1, \ldots, A_n$ to B, namely the following:

(1) If A_1 and . . . and A_n, then B

is valid.

In modern logic, forms are represented by formulas, the notion of validity is applied in the first instance to formulas, and validity of a conclusion is identified with being an *instance* of a valid formula: the result of substituting specific sentences, predicates, or whatever for the sentence letters, predicate letters, or whatever in the formula (and if quantifiers are involved, fixing what they are supposed to range over). Intuitively, a formula is supposed to be valid if it is *guaranteed* that all instances *must* be true, but

logicians could never begin work on classifying formulas as valid or invalid if they had to wait for philosophers of logic to achieve agreement as to just what the "guarantee" or "must" here amounts to. For present purposes we need only a definition that gets the extension right, and may take a formula to be valid if all instances *are* true, never mind about *must*. Such a definition might not get the extension right if our language were severely impoverished, but fortunately it isn't. Or at least, it presumably isn't on the understanding, here adopted, that instantiation may be by expressions containing free variables (or analogous expressions of natural language), with truth in all instances requiring truth of all instantiations *for all values of any free variables (or analogous expressions)*. For example, *Px* may be instantiated by "*x* is an element of the set *u*," with validity requiring truth *whatever set the u in question is taken to be*, or by "*x* likes it," with validity requiring truth *whatever item the it in question is taken to be*.

One can show that a given conclusion is *not* a consequence or implication of given premises by replacing the nonlogical words in premises and conclusions by letters to give a formula, and then presenting an instantiation in which other nonlogical words to replace the letters in which the premises become true and the conclusion false. This is not essentially different from the method traditionally used in mathematics to show that, for instance, the parallel postulate does not follow from other axioms of Euclidean geometry, except that traditionally the form was not explicitly exhibited, and one simply presented a substitution of new nonlogical words for the original ones.

The most basic goals of model theory for pure logic are to define rigorously: first, a technical notion of model *capturing all and only the features of an instance that are relevant to truth*; and second, a technical notion of truth in a model such that *truth in the intuitive sense of an instance will coincide with truth in the technical sense in a corresponding model*. It then follows that the intuitive notion of validity as truth in all instances will coincide with the rigorously defined notion of validity as truth in all models. For this coincidence to obtain and be useful, the class of models must be (a) broad enough that there is a model corresponding to any instance; and (b) narrow enough to be more tractable than the class of all instances.

At the level of classical sentential logic, instantiation will simply be substitution of sentences $\pi, \pi', \ldots$ for sentence letters $p, p', \ldots$, and since all that matters for the truth of the result is the truth-values of the sentences, and not, for instance, their meaning, for a *model* we may simply take an assignment of truth-values to sentence letters. We can then define the notion $\mathbf{M} \models A$ of the truth in a model $\mathbf{M}$ of a formula A built up from sentence letters using negation and conjunction and disjunction $\sim$ and $\wedge$ and $\vee$, inductively as follows:

(2a)	$\mathbf{M} \models p_i$	iff	$\mathbf{M}(p_i) = \text{true}$
(2b)	$\mathbf{M} \models \sim A$	iff	not $\mathbf{M} \models A$
(2c)	$\mathbf{M} \models A \wedge B$	iff	$\mathbf{M} \models A$ and $\mathbf{M} \models B$
(2d)	$\mathbf{M} \models A \vee B$	iff	$\mathbf{M} \models A$ or $\mathbf{M} \models B$

(Similarly for the conditional $\rightarrow$ and biconditional $\leftrightarrow$.) A valid formula is then one that is true in all such models. The desiderata (a) and (b) are achieved, the former obviously, the latter outstandingly, since for a formula containing only specific sentence letters (for example, the three p, q, r) there will be only finitely many (in the example, just eight) models or combinations of truth values, whereas there are an infinity of instances (triples of sentences that might be put in for p, q, r). This reduction from infinite to finite is responsible for the decidability of classical sentential logic.

At the level of classical predicate logic, the notion of model becomes more complicated. What matters for the truth of an instance is, first of all, which things are being spoken of, or in jargon, the *universe of discourse* or *quantifier domain*; and further, for each *k*-place predicate, which *k*-tuples of elements of that universe it is true of. (If constants are present, they must be assigned distinguished elements of the domain; and if function symbols are present, they must be assigned functions on the domain, of the right number of places; but for simplicity we ignore such complications.) Differences in *meaning* between the predicates in different instantiations will not matter, so long as the predicates are *true* of the same things, so the notion of model may ignore them. The official definition of a model **M** accordingly takes one to consist of a nonempty set M as quantifier domain, and an assignment to each *k*-place predicate letter *P* of a set $P^{\mathbf{M}}$ of *k*-tuples of elements thereof, called **M**'s *interpretation* of *P*. The notion of truth in a model also becomes more complicated, involving a detour through the notion of an *open* formula's *satisfaction* by elements in a model; but Tarski's definition is given in all standard logic textbooks today, and need not be repeated here.

As for desideratum (b), the class of models is vast, but basic results of model theory imply that truth in all models is equivalent to truth in all models of a restricted kind. To begin with, the *Löwenheim-Skolem* theorem states that a formula will be true in all models iff it is true in all models with a countable quantifier domain.

As for desideratum (a), there is a problem. We want to apply classical logic to evaluate arguments in, among other fields, set theory. We need therefore to consider formulas with a single two-place predicate letter and the instantiation in which the quantifier domain consists of all sets, and what is put in for the predicate letter is "is an element of." But there is no model corresponding to this instantiation, since our official definition of model requires the quantifier domain to be a set, and there is no set of all sets. In Kreisel (1967), where the problem was first noted, a solution is provided, based on comparison of the notions of: (i) formal provability, (ii) validity in the intuitive sense (truth in all instances, even if the domain is not a set), (iii) validity in the official sense (truth in all models, where the domain must be a set). Intuitively, (i) is found to imply (ii) on inspecting the axioms and rules of formal proof; and obviously (ii) implies (iii); but the Gödel completeness theorems tells us (iii) implies (i), so the intuitive notion (ii) is caught in a scissors between the technical notions (i) and (iii), and all three notions agree. (For second-order logic, by contrast, and with it plural logic, there is no completeness theorem and no scissors, and the assumption that every formula that is valid in the official sense is valid in the intuitive sense is known to be unprovable from the usual axioms of set theory, leaving the ultimate significance of the model theory in some doubt.)

Tense logic

The simplest time-distinctions are expressed in English not through overt quantification over "times" but by verb inflections or auxiliary verbs. There is nothing comparable to these verb modifications in classical logic, which was designed for application to pure mathematics, where nothing ever has been or ever will be other than as it is. Tense logic as developed in Prior (1967) and elsewhere introduces operators F and P, pronounced "(sometime) will be" and "(sometime) was." In terms of these, other operators G = ~F~ and H = ~P~, pronounced "*always* will be" and "*always* was" can be defined. For simplicity we consider only the future pair F, G in examples. The form of such a natural language sentence as

(3) If he ever goes, she will go later

might be represented using these operators thus

(4) $G(p \rightarrow Fq)$

with *p*/*q* for "he/she goes."

Though nothing ever changes in mathematics, mathematics nonetheless can be and is applied to physics, where everything changes. The application requires some departure from our usual ways of speaking and thinking, adding a phrase "at time *t*" to each verb, and replacing the present tense "is" that contrasts with "was" and "will be" with a tenseless "|is|" that is short for "was or is or will be." Taking this approach, the form of the premise (3) might be represented classically as

(4') $\forall t(t_0 < t \wedge Pt \rightarrow \exists t'(t < t' \wedge Qt'))$

with t_0 standing for the present time, < for the relative-futurity or earlier-later relation, and *Pt* or *Qt* for "he or she |goes| at time *t*." Comparison of (4) and (4') suggests how in general tense-logical formulas can be transcribed as classical-logical formulas, hence how any natural language representable by a tense-logical formula is representable by a classical-logic formula. Each G or F will correspond to an $\forall$ or $\exists$, each sentence letter p_i will correspond to a one-place predicate letter P_i, with $P_i t$ to be thought of as meaning something like "p_i |is| true at time *t*."

Classical first-order logic counts the premises as implying the conclusion in very few tense-logical arguments as they stand, though the conclusion is often a consequence of the stated premises plus some further unstated premise about the structure of time, such as the transitivity of the earlier-later relation:

(5) $\forall t \forall t' \forall t''(t < t' \wedge t' < t'' \rightarrow t < t'')$

It is the job of the physicist to say which such assumptions are true, and of the logician to say what follows from each. Tense logic seeks to develop systems of axioms and rules for deriving tense-logical theorems working entirely within the tense-logical language. The tense logician would like to be able to tell the physicist which axioms hold on which assumptions about time, but there is this difficulty, that the physicist's assumptions about time are formulated in the mathematical style (5) as principles about a timeless relation of earlier to later among entities called times, and the kind of language with which tense logicians directly work. It is in order to connect the two languages that tense logicians introduce a notion of model that is a direct adaptation of the classical notion of model for the first-order language with one two-place predicate < and many one-place predicates P_i.

A model **M** consists of (i) a set M of elements representing "times," (ii) a two-place relation $<^{\mathbf{M}}$ on the set representing "earlier-later," and (iii) a component telling us for each sentence letter p_i and each time *t* whether p_i is true or not at *t*. Here (i) and (ii) are just are the classical quantifier domain and interpretation of <. Classically (iii) would be represented by an assignment to each P_i of an interpretation $P_i^{\mathbf{M}}$, to be thought of as the set of times when p_i is true; tense-logically, (iii) is more conveniently represented by a function assigning each time *t* a function $\mathbf{M}_t$ assigning each sentence letter a truth value; the two representations are mathematically equivalent.

The notion of the truth of a formula A at a time t in a model $\mathbf{M}$, written $\mathbf{M}, t \models A$ can then be defined by recursion on the complexity of the formula, adapting the classical definition thus:
DISPLAY

(6a)	$\mathbf{M}, t \models p_i$	iff	$\mathbf{M}_t(p_i) = \text{true}$
(6b)	$\mathbf{M}, t \models \sim A$	iff	not $\mathbf{M}, t \models A$
(6c)	$\mathbf{M}, t \models A \wedge B$	iff	$\mathbf{M}, t \models A$ and $\mathbf{M}, t \models B$
(6d)	$\mathbf{M}, t \models A \vee B$	iff	$\mathbf{M}, t \models A$ or $\mathbf{M}, t \models B$
(6e)	$\mathbf{M}, t \models \mathrm{F}A$	iff	$\mathbf{M}, t' \models A$ for some t' with $t <^{\mathbf{M}} t'$
(6f)	$\mathbf{M}, t \models \mathrm{G}A$	iff	$\mathbf{M}, t' \models A$ for all t' with $t <^{\mathbf{M}} t'$

A model may be required to have also (iv) a designated element t_0 representing the "present," in which case truth in the model as a whole is defined as truth at t_0, and validity as truth in all models; more usually, validity is simply defined as truth at all times in all models.

A certain minimal axiom system has been identified, which gives as theorems all formulas that are valid without special assumptions about the structure of time, and various additional axioms corresponding to various such special assumptions have also been identified. For instance, corresponding to transitivity (5) we have the axiom

(7) $\mathrm{FF}p \rightarrow \mathrm{F}p$

intuitively meaning roughly

(8) Whatever *will be* going to be *is* going to be.

Prior offered something like (8) as the "cash value" of what is asserted using tenseless mathematical language by something like (5). The "cash values" of quite sophisticated physical theories of time, as in relativity theory, have been worked out by tense logicians.

If one confused "semantics" in the sense of model theory with "semantics" in the sense of meaning theory, one would likely be led to conclude that it is not (8) that gives the cash value of (5), but (5) that gives the cash value of (8). One would likely be led, that is to say, to the conclusion that sentences involving tense-operators—and presumably therewith the sentences of tensed natural language whose surface features tense-operators were intended to represent—deep down really *mean* something that quantifies over such entities as "times" or "instants": that what appears to be an historically late, sophisticated, philosophico-scientific posit, the "instant of time," was really an implicit assumption of ordinary speech all along.

But any conclusion of this sort about natural language would be an empirical linguistic hypothesis, requiring empirical linguistic evidence. There may well be good evidence of the right kind for some such hypothesis; but if so, that only makes it the more important for the student of these matters to be alert to the distinction between such genuine empirical linguistic evidence and the spurious support that would be provided by an equivocating argument sliding fallaciously from the presence of "times" or "instants" in models to their presence in "semantics" to their presence in meaning.

Modal logic

Modal logic is concerned with the distinction between what actually is and what is in one sense or another necessary or possible. The simplest modal distinctions are expressed in English through modal auxiliary verbs such as "must" and "may." There is nothing comparable in classical logic, because in pure mathematics, for which that logic was designed, nothing could have been other than as it is. Modal logic as developed by C. I. Lewis (1918) and his successors introduces operators $\Box$ and $\Diamond$, pronounced "necessarily" and "possibly," the two being interdefinable since $\Box = {\sim}\Diamond{\sim}$ and $\Diamond = {\sim}\Box{\sim}$, and develops systems of axioms and rules for deriving modal theorems.

Historically, systems quickly proliferated without much regard to intuitive interpretation, differing in their axioms regarding iterated modalities, such as

(9a) $\Box p \rightarrow \Box\Box p$
(9b) $\Diamond p \rightarrow \Box\Diamond p$

For instance, the systems known as **S4** and **S5** both have (9a) as a theorem, while the former lacks but the latter has (9b) as well. Modal logic was pursued for half a century, and dozens of systems were developed, before logicians gave much attention even to so basic a distinction as that between *metaphysical* possibility (concerning what potentially might have been if the world had been otherwise than it actually is) and *logical* possibility (concerning what can without internal self-contradiction be supposed about how the world actually is).

The fact that formal development ran ahead of intuitive interpretation was in some way a good thing, since it left the formalism open to unanticipated interpretations and applications, which have included the *epistemic interpretation of intuitionistic logic*, as in Shapiro (1985), the provability interpretation of modal logic, or *provability logic* for short, as in Boolos (1995), as well as what is called *dynamic logic* in theoretical computer science. But it does leave the logician with the question which of the many systems in existence is appropriate for which of the various senses of modality that have been distinguished.

During the period before logical and metaphysical modality were clearly distinguished, a model theory was developed for many modal systems, primarily Kripke (1963a), in order to address such issues as decidability. Successively considering such formulas as

(10a) ${\sim}(p \wedge q) \wedge \Diamond(p \wedge q)$
(10b) ${\sim}(p \wedge q) \wedge {\sim}\Diamond(p \wedge q) \wedge \Diamond\Diamond(p \wedge q)$
(10c) $\Diamond(p \wedge q \wedge \Diamond({\sim}p \wedge q)) \wedge \Diamond(p \wedge {\sim}q \wedge {\sim}\Diamond({\sim}p \wedge q))$
(10d) $\Diamond(p \wedge q \wedge \Diamond({\sim}p \wedge q)) \wedge \Diamond(p \wedge q \wedge {\sim}\Diamond({\sim}p \wedge q))$

will lead to successively concluding that a model will have to (a) represent not only what combination of truth values the sentences used to instantiate sentence letters actually have, but also what combinations they could possibly have had, and (b) represent possibilities "of different orders," not only actual possibilities but also possible possibilities, and (c) represent not only possibilities of various orders but also which higher-order possibilities are possible relative to which lower-order possibilities, (d) allow multiple possibilities for which the sentences involved have the same combination of truth values.

All these features are present in a *Kripke model* **M**, which consists of (i) a set M of elements representing possibilities of various orders, (ii) a two-place relation $<^{\mathbf{M}}$ on the set representing relative possibility, and (iii) a function assigning each element t of the set a function $\mathbf{M}_t$ assigning each sentence letter a truth value. This is, of course, exactly what a model for tense logic amounts to, and the same definition of truth in a model is used in both logics, $\Box$ and $\Diamond$ replacing G and F, so that clauses (6abcd) are retained, and (6ef) replaced by

(11e) $\mathbf{M}, t \models \Diamond A$ iff $\mathbf{M}, t' \models A$ for some t' with $t <^{\mathbf{M}} t'$
(11f) $\mathbf{M}, t \models \Box A$ iff $\mathbf{M}, t' \models A$ for all t' with $t <^{\mathbf{M}} t'$

(Historically, the models for modal logic came first, and those for tense logic later.) Traditionally, the elements of the set M are called "worlds" and the relation $<^{\mathbf{M}}$ "accessibility."

A certain minimal axiom system **K** has been identified that gives as theorems all formulas that are valid without special assumptions about the accessibility relation among worlds; and various additional axioms corresponding to various such special assumptions have also been identified. Notably

(12a) $\Box p \rightarrow p$
(12b) $\Box p \rightarrow \Box\Box p$
(12c) $p \rightarrow \Box\Diamond p$

correspond respectively to the assumptions of reflexivity and transitivity and symmetry of accessibility.

But does this tell us anything about which of the many modal axioms systems is appropriate for which of the various notions of necessity that have been distinguished? Here the contrast with tense logic is striking. Physicists' assumptions about the structure of time do naturally present themselves as assumptions about the earlier-later relation among times. Different notions of necessity do *not* naturally present themselves as assumptions about the "accessibility" relation among "worlds."

This is not to say that models are useless, or useful only for technical purposes. About a decade elapsed between the time when the model theory became well known through Kripke (1963b) and the time when the distinction between metaphysical and logical modality became well known through Kripke (1972), but during that decade the model theory already had an important influence even on intuitive as opposed to technical issues. It motivated the general rejection of the so-called *converse Barcan formula*

(13) $\Box\forall x Px \rightarrow \forall x \Box Px$

by confirming preexisting suspicions that this formula somehow was saying that whatever exists necessarily exists, and also by suggesting an axiom system that avoids (13) as a theorem. (On the other hand, the mere existence of a formal model theory, which in itself shows nothing more than that the axioms systems are formally consistent, encouraged complacency in the face of Quine's questioning whether "quantifying in" to modal contexts made any intuitive sense. If metaphysical modality is meant, it does; if logical modality is meant, it does not, or so it is argued in Burgess (1988).)

Moreover, even in connection with the question of which axiom system is appropriate for which notion of modality, the model theory has played an auxiliary role. If one wishes to argue that **S5** is *sound*—that every theorem of the system is true in all instances—when the box is read as "it is true by virtue of logical form that," there is no reason to consider models. One may simply argue intuitively that whatever sentence π is put in for p in (12c), the result will be true, and similarly for every other axiom.[9] It is when one wishes to argue that **S5** is *complete* for logical necessity in the indicated sense—that every formula true in all instances is a theorem, or equivalently, that every formula whose negation is not a theorem is true in some instance—that the model theory becomes useful.

To illustrate how, consider a formula whose negation is not a theorem, such as

(14) $\Diamond(p_1 \wedge p_2) \wedge \Diamond(p_2 \wedge p_3) \wedge \Diamond(p_3 \wedge p_1) \wedge \sim\Diamond(p_1 \wedge p_2 \wedge p_3)$

What we would like to show is that it has an instantiation that is true. What model theory tells is is that it has a model in which it is true. What we need to do is derive an instantiation from a model.

To show how this may be done, consider the model with three worlds w_1, w_2, w_3, all accessible from each other, and with each p_i true at precisely the w_j for $j \neq i$. We first want, corresponding to the three worlds w_i of the model, three sentences ω_i whose logical form will guarantee that exactly one of them is true. These can be obtained starting from any two logically independent sentences σ_1 and σ_2, say:

(15a) σ_1 = "Snow is white"
(15b) σ_2 = "Grass is green"

as follows:

(15c) ω_1 = σ_1 and σ_2
(15d) ω_2 = σ_1 and not σ_2
(15e) ω_3 = not σ_1

(For a k-world model, we would need to start with n independent sentences, where $k \leq 2^n$.) Then to each p_i, associate the disjunction π_i of the ω_j corresponding to w_j where it is true. The results amount to

(15f) π_1 = "Snow is not white or grass is not green"
(15g) π_2 = "Snow is not white or grass is green"
(15h) π_3 = "Snow is white"

Logical form does not preclude any two of the π_i being true but does preclude all three being true, as required to give a true instance of (14). The method generalizes.

The correct logic for the provability interpretation has also been determined, and again models play an auxiliary role, though most of the hard work is getting from a model for a formula to a true instance of that formula. There is not space to enter into details here, nor to enter into the situation in other nonclassical logics, such as conditional or intuitionistic or relevance/relevant logic. In every case, mathematical results about model theory can be useful for philosophical purposes when properly employed.

In every case, the model theory does not come with directions for its proper employment, which rather can only be discovered by philosophical analysis and reflection.

References

Boolos, George (1995) *The Logic of Provability*, Cambridge: Cambridge University Press.

Burgess, John P. (1998) "Quinus ab omne nævo vindicatus," in A. A. Kazmi, ed., *Meaning and Reference*, *Canadian Journal of Philosophy Supplement* vol. 23, 25–65.

Kreisel, Georg (1967) "Informal rigor and completeness proofs," in I. Lakatos, ed., *Problems in Philosophy of Mathematics*, Amsterdam: North Holland, 138–186.

Kripke, Saul (1963a) "Semantical analysis of modal logic 1, normal propositional calculi," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, vol. 9, 113–116.

(1963b) "Semantical considerations on modal logic," *Acta Philosophical Fennica*, vol. 16, 83–94.

(1972) "Naming and necessity," in D. Davidson & G. Harman, eds., *Semantics of Natural Language*, Dordrecht: Reidel, 253–355 and 763–769.

Lepore, Ernest & Kirk Ludwig (2007) *Donald Davidson's Truth-Theoretic Semantics*, Oxford: Oxford University Press.

Lewis, Clarence Irving (1918) *A Survey of Symbolic Logic*, Berkeley: University of California Press.

Morris, Charles (1938) *Foundation of the Theory of Signs*, International Encyclopedia of Unified Science, vol. 1, no. 2, Chicago: University of Chicago Press.

Prior, Arthur (1967) *Past, Present and Future*, Oxford: Clarendon Press.

Shapiro, Stewart (1985) (ed.) *Intensional Mathematics*, Amsterdam: North Holland.

5.2
LOGICAL QUANTIFIERS

Gila Sher

This chapter offers a logical, linguistic, and philosophical account of modern quantification theory. Contrasting the standard approach to quantifiers (according to which logical quantifiers are defined by enumeration) with the generalized approach (according to which quantifiers are defined systematically), the chapter begins with a brief history of standard quantifier theory and identifies some of its logical, linguistic, and philosophical strengths and weaknesses. It then proceeds to a brief history of generalized quantifier theory and explains how it overcomes the weaknesses of the standard theory. One of the main philosophical advantages of the generalized theory is its philosophically informative criterion of logicality. The chapter describes the work done so far in this theory, highlights some of its central logical results, offers an overview of its main linguistic contributions, and discusses its philosophical significance.

I Standard Modern Quantifier Theory

1 Historical Beginnings

Modern quantifiers were first introduced by Frege (1879). The quantifiers All and Some were already recognized as logical operators by Frege's predecessors, going all the way back to Aristotle. But Frege offered a new analysis of these operators that enabled him to deal with multiple-quantifier prefixes and, more generally, multiple nesting of quantifiers. This, in turn, enabled him to extend traditional logic from a logic of "terms" (1-place predicates standing for properties) to a logic of relations. These changes revolutionized logic. Among the many achievements of Frege's theory, three are especially relevant for the present discussion:

(a) Frege's theory generated a considerably larger array of quantifiers than traditional logic: Starting with All as his basic logical quantifier, Frege construed not just the traditional Some, No, and Not All as (defined) logical quantifiers but also infinitely many others, for example, Exactly n, At Least n, and At Most n, for every natural number n.
(b) Frege's theory offered a systematic interpretation of multiple-quantifier prefixes and other nesting quantifiers.
(c) Frege's expansion of logical quantifiers gave rise to languages with enormous expressive power. Extensive parts of natural language as well as the entire

language of classical mathematics and many segments of the language of science are expressible using his quantifiers.

Frege regarded 1st-order quantifiers as 2nd-order functions or concepts. The 1st-order quantifier Some is the 2nd-order concept of nonemptiness, the 1st-order All is the 2nd-order concept of universality. Using contemporary terminology, "$(\exists x)\Phi x$" says that the extension of Φ is not empty, and "$(\forall x)\Phi x$" says that the extension of Φ is universal.

Frege sanctioned both 1st-order and higher-order quantifiers. The former quantify over individual variables, i.e., variables ranging over individuals; the latter quantify over functional and predicative variables, i.e., variables ranging over functions, properties, and relations. For reasons that go beyond the scope of the present chapter (but see discussion of completeness in Section II.1 below), most logicians, linguists, and philosophers today focus on 1st-order quantifiers. In what follows I will limit myself to such quantifiers.

Note: Throughout this chapter I am using contemporary terminology.

2 Semantic and Proof-Theoretic Definitions of Quantifiers

Tarski (1933) developed a semantic definition of truth for languages formulated within the framework of modern logic. The recursive entries for the standard quantifiers—All ($\forall$) and Some ($\exists$)—are: given an assignment g of objects (in the intended universe) to the variables of the language,

($\exists$) "$(\exists x)\Phi x$" is true under g iff (if and only if) at least one object *a* (in the universe) satisfies "Φx" under g;
($\forall$) "$(\forall x)\Phi x$" is true under g iff every object *a* (in the universe) satisfies "Φx" under g;

where *a* satisfies "Φx" under g iff it satisfies "Φx" when all the free variables of "Φx" other than x are assigned values by g.

Gentzen (1935) developed a proof-theoretic method for defining logical constants in terms of *introduction* and *elimination rules*. Such rules tell us, for a given logical constant, under what conditions we can derive a statement governed by it from a statement (or statements) not governed by it, and under what conditions we can derive a statement (or statements) not governed by it from one governed by it. The entries for the standard logical quantifiers are, essentially:

($\exists$) *Introduction:* $\Phi(a) \vdash (\exists x)\, \Phi x$
Elimination: $(\exists x)\, \Phi x \vdash \Phi(a)$ provided "a" plays a merely auxiliary role
($\forall$) *Introduction:* $\Phi(a) \vdash (\forall x)\, \Phi x$ provided "a" is arbitrary
Elimination: $(\forall x)\, \Phi x \vdash \Phi(a)$

where "a" is an individual constant and certain syntactic conditions are given for " 'a' plays a merely auxiliary role" and " 'a' is arbitrary." Frege's, Tarski's, and Gentzen's characterizations identify roughly the same quantifiers, though the last generalizes differently from the first two.

II Strengths and Weaknesses of Standard Quantifier Theory

1 Logical Results

The logic of the standard quantifiers, i.e., standard 1st-order logic, is rich in logical, or rather meta-logical, results. Three results that are especially relevant for our discussion are:

(a) *The Löwenheim-Skolem Theorem* (1915/20): If a sentence σ (set of sentences Σ) has a model, it has a countable model, i.e., a model whose cardinality (i.e., the cardinality of its universe) is smaller than or equal to the cardinality of the natural numbers ($\aleph_0$).

(By definition, σ/Σ has a model iff there is a model M such that σ/all the sentences of Σ is/are true in M.) This result is naturally viewed as a limitation on the expressive power of standard 1st-order logic: Within the framework of this logic we cannot "seriously" define the notion of "uncountably many" and many other infinitistic notions, since every consistent statement of the form "There are uncountably many F's" comes out true in some countable model, i.e., a model in which there are only countably many F's. This result is sometimes called "Skolem's Paradox," although it is not a real paradox. But it does imply that the representational powers of the standard 1st-order quantificational languages are limited, since they cannot determine uniquely the intended meanings of formal terms and, as a result, the meaning of claims (or theories) containing such terms. For example, a consistent theory that says there are uncountably many objects has a model (i.e., is true in a model) that does *not* have uncountably many objects.

(b) *The Completeness Theorem* (Gödel 1930): Standard 1st-order logic has a complete proof system. I.e., A sentence σ follows logically from a set of sentences Σ in the semantic sense ($\Sigma \models \sigma$) iff it follows logically from it in the proof-theoretic sense ($\Sigma \vdash \sigma$).

The completeness result is perhaps the most important theoretical result of standard quantification theory. It says that in standard 1st-order logic the proof-theoretic relation of logical consequence is just as powerful as the semantic relation of logical consequence. This result is one of the main reasons most contemporary philosophers and logicians prefer standard 1st-order logic to standard (full) 2nd-order logic. (Full) 2nd-order logic is incomplete: Its proof system is less powerful than its semantic system.

One consequence of the completeness of standard 1st-order logic is the co-extensionality of the semantic and proof-theoretic definitions of the standard quantifiers. Since proofs, unlike models, are finite structures, this might be viewed as an advantage of the proof-theoretic definition over the semantic definition.

(c) *The Compactness Theorem* (Gödel 1930). Two significant formulations are: (i) If σ follows logically from an infinite set Σ, then it follows logically from a finite subset of Σ; (ii) If every finite subset of Σ has a model, Σ has a model.

The compactness result is naturally viewed as an advantage: it enables us to deal with infinite sets by dealing with their finite subsets.

It should be noted, however, that the connection of these theorems to the standard 1st-order quantifiers is not always simple. For example, compactness and completeness fail for logics like $L_{\infty,\omega}$, which has formulas with arbitrarily large conjunctions and disjunctions, but finitely many quantifiers.

2 Linguistic Strengths and Weaknesses

Standard quantifier theory made a substantial contribution to linguistic semantics and syntax through the works of Geach (1962), Montague (1973), Kamp (1981), May (1985), Partee (1987), and numerous others. Not only did it allow linguists to analyze complex linguistic structures involving multiple-quantifier prefixes (e.g., "$(\forall x)(\exists y)$," "$(\exists x)(\forall y)(\exists z)$") and other nesting quantifier structures (e.g.,"$(\forall x)[(\exists y) \ldots \supset (\exists y)$—]"), it enabled them to formulate important questions concerning quantifier scope, indefinite nouns, anaphora, etc. A well-known problem of this kind is the quantificational structure of "donkey sentences," i.e., sentences like

(1) Every farmer who owns a donkey, beats it.

(See discussion in, e.g., Neale 1990.) Moreover, the theory of standard quantifiers played a central role in the development of important linguistic theories (for example, LF theory), the establishment of new semantic frameworks (for example, Montague grammar), and so on.

But the expressive power of the standard quantifiers is linguistically very limited. Even such simple quantitative determiners as "most", "few", "half", "an even number of", "all but ten", "finitely many", "uncountably many", and so on, cannot be expressed by any (well-formed) combination of standard 1st-order quantifiers. As far as quantifier-structures in natural language are concerned, it is quite clear that standard quantifier theory offers no more than a preliminary framework for their study.

3 Philosophical Strengths and Weaknesses

Standard quantifier theory has had a privileged status in contemporary analytic philosophy since its birth in the early years of the 20th century. It is interwoven into most analytic theories, and its position has been sealed by such works as Quine's influential *Philosophy of Logic* (1970). Even philosophers who have challenged the standard theory have rarely viewed its choice of quantifiers as a problem. They have assigned $\forall$ and $\exists$ a large array of alternative interpretations, going all the way from the substitutional interpretation (Marcus 1972) to the intuitionistic (Dummett 1973) and paraconsistent (Priest & Routley 1989) interpretations, but few have contested its restriction to the standard quantifiers. Perhaps the most celebrated feature of the standard theory is the *completeness* of its proof method, i.e., the fact that it is integrated into a powerful logic with a complete proof system (see above). This, indeed, is a desirable feature. For many, including Quine, the incompleteness of stronger quantifier theories justifies their rejection.

But standard quantifier theory is also widely recognized to have many problems. Here I will focus on six problems of philosophical significance (some mentioned above, and some interrelated):

(a) *Inability to Account for Basic Inferences*

In his 1962 abstract, Rescher points out that standard "quantificational logic is unequal to certain childishly simple valid arguments" like "Most things are A's, Most things are B's; therefore: Some A's are B's", and "Most C's are A's, Most C's are B's; therefore: Some A's are B's" (ibid.: 374). Since these inferences are of the same kind as the inferences logic is designed to account for (for example," All things are A's, Some things are B's; therefore: Some A's are B's"), this state of affairs is problematic.

(b) *Limited Logical Expressive Power*

As noted in Section II.2 above, many elementary quantificational notions, including quantitative determiner notions like "most", "few", "half", "finitely many", and "uncountably many" are indefinable in terms of the standard quantifiers (and other standard logical constants).

(c) *Limited Formal Expressive Power*

As demonstrated by the Löwenheim-Skolem Theorem (Section II.1(a) above), many formal (mathematical) notions, and in particular infinitistic notions like "uncountable", are not only indefinable as logical notions within the standard quantificational framework but are also inadequately definable as nonlogical notions within that framework. That is, it appears that the standard framework does not enable us to define them in a way that adequately captures their meaning. (In logicians' terminology, it enables us to define "uncountable" only as a "relative" notion but not as an "absolute" notion.)

(d) *Absence of a Theoretical Criterion for Logical Quantifiers*

It is a surprising fact about modern logic that it has a theoretical, precise, systematic, informative, and philosophically explanatory criterion for logical *connectives* but not for logical *quantifiers* or predicates. Logical connectives (at least in classical logic) have a precise and philosophically informative criterion, *truth-functionality*, which is given a mathematically precise definition by Boolean algebra. Every logical connective is represented by a Boolean operator in a 2-element Boolean algebra, and every Boolean operator of that kind determines a logical connective. Philosophically, every truth-functional connective—i.e., a connective that takes into account only the formal pattern of truth-values of its arguments—is logical, and every logical connective is truth-functional. (On the philosophical ramifications of this criterion, see (e) below.)

In contrast, for logical quantifiers and predicates we have nothing more than an utterly uninformative, unexplanatory, and unsystematic characterization, namely, a characterization by *enumeration*. All this characterization says is that $\forall$ and $\exists$ are logical quantifiers and $=$ is a logical predicate, and that any quantifier definable from $\forall$ and $\exists$ (using logical predicates and connectives) is also a logical quantifier, and any predicate defined from $=$ (using other logical constants) is also a logical predicate. Why *these rather than others* are logical, why *only these* are logical, remains a mystery. We may try to justify the standard choice on pragmatic grounds, by appealing to custom or convenient features of the associated logic (for example, completeness), but unless we add a theoretical dimension to such a justification it remains philosophically shallow. Furthermore, even

on purely pragmatic grounds, how do we know that a stronger 1st-order system, i.e., a 1st-order system with additional logical quantifiers, will not give us an overall more desirable logic? The current situation is thus philosophically unacceptable.

(e) *Lack of Philosophical Understanding of Logicality*

Given the central role that logic plays in all areas of knowledge, including philosophy, the nature of logicality requires a theoretical account. The nature of logicality, however, has so far resisted such an account. One of the problems facing us is finding a fruitful standpoint from which to launch the investigation, i.e., a standpoint, or a perspective, which will enable us to formulate the question in a precise, systematic, and philosophically enlightening terms. It is quite clear that such a perspective requires an informative account of the logical constants. We know that logical truth, logical consequence, logical consistency, etc., are largely due to the logical constants involved; understanding the nature of the logical constants is therefore a prerequisite to a theoretical understanding of logicality.

Our experience with sentential logic shows that a theoretical criterion for logical constants can, indeed, make a significant contribution to our philosophical understanding of logicality. Philosophically, the truth-functionality of the logical connectives means that they *abstract* from everything but the bare *truth* pattern of their arguments, and this explains why logical consequences based on these connectives preserve *truth* rather than other features of sentences. Furthermore, this, together with the fact that truth-functional connectives correspond to mathematical operators, i.e., operators governed by *necessary* laws, and the fact that the test for logical truth/consequence based on such connectives is truth/preservation-of-truth under *all possible* variations in truth patterns, explain some of the philosophically significant features of sentential logic. In particular, they explain its great generality, topic neutrality, immunity to (at least most) empirical discoveries, and necessity. A philosophically informative criterion for *logical quantifiers* (and *predicates*) is needed to extend this understanding to all of logic.

(f) *The Shaky Status of the Semantic Definition of Logical Consequence for Quantificational Logic*.

In his 1936 paper, "On the Concept of Logical Consequence," Tarski constructed a semantic definition of logical consequence for quantificational logic that established logical semantics (model theory) as one of the two major fields of contemporary logic, along with proof theory. Tarski's semantic definition says that:

(LC) $\Sigma \models \sigma$ iff every model of Σ is a model of σ (or: there is no model in which all the sentences of Σ are true and σ is false).

Tarski anchored his semantic definition in the intuitive idea that *a logical consequence is a necessary-and-formal consequence*. To determine the adequacy of his definition, he asked whether it accurately captured this idea. Investigating this question, he realized that the answer varies according to the choice of logical constants. For some choices of logical constants—for example, treating Is-a-Property-of-Napoleon as a logical quantifier or treating $\forall$ as a nonlogical constant—the answer is clearly negative, while for others, including the standard choice of logical constants, no conflict seems to occur.

But even in the latter case, Tarski realized, it is not clear that his definition fully captures the idea it was supposed to capture. Adding such nonstandard logical quantifiers as Finitely Many also accords with this idea. The scope and justification of the contemporary definition of logical consequence, thus, require a systematic criterion for logical quantifiers.

A solution to all the above problems is offered by *generalized quantifier theory*.

III Generalized Quantifier Theory

1 Brief History

The history of generalized quantifiers starts with Mostowski (1957). Mostowski suggested that we might generalize the standard notion of logical quantifier along two dimensions, syntactic and semantic. Syntactically, a logical quantifier is a variable-binding operator that generates new formulas from old formulas. Semantically, a logical quantifier over a universe A is a cardinality function from subsets of A to a truth-value, satisfying a certain invariance condition. For example, the existential and universal quantifiers over A, $\exists_A$ and $\forall_A$, are functions that, given a subset B of A:

(2) $\exists_A(B) = T$ iff the cardinality of B, $|B|$, is larger than 0; i.e., iff $|B| > 0$.
(3) $\forall_A(B) = T$ iff the cardinality of the complement of B in A is 0; i.e., iff $|A–B| = 0$.

It is easy to see that both $\exists$ and $\forall$ satisfy the following invariance condition:

(4) A quantifier Q on A is invariant under all permutations of A.

That is: For every A, B, B', where A is a nonempty set and B, B' are subsets of A: If B' is obtained from B by some permutation of A, then: $Q_A(B) = Q_A(B')$, where a permutation of A is a bijection p: $A \rightarrow A$, i.e., a 1–1 function from A onto A. Using "$\subseteq$" for "is a subset of" and "p*" for "the function on subsets of A induced by p", we can spell out (4) by:

(5) If A is a nonempty set, $B \subseteq A$, $B' \subseteq A$, and there is a permutation p of A such that p*(B) = B', then: $Q_A(B) = Q_A(B')$.

Generalizing, Mostowski arrived at the following criterion for logical quantifiers:

(PER) A Quantifier Q is logical iff for any non-empty set A, Q_A is invariant under all permutations of A.

Under this criterion, every 2nd-order 1-place cardinality predicate is a 1st-order logical quantifier, and every 1st-order logical quantifier is a 2nd-order 1-place cardinality predicate. Examples of logical quantifiers in addition to $\exists$ and $\forall$, are Most (interpreted as, say, "more than half"), An Even Number of, Finitely Many, Uncountably Many, and α-Many for every cardinal number α, finite or infinite. Such quantifiers appear in English sentences such as

(6) Most things are beautiful.
(7) There is a finite number of stars.

They are symbolized by the formulas

(8) (Mx)Bx
(9) (Fx)Bx,

using obvious abbreviations.

Examples of 2nd-order 1-place predicates that are not logical quantifiers are "some men", "all women", "is a property of Napoleon", and "is a color property".

Lindström (1966) generalized Mostowski's criterion further in two ways. First, he extended it to 1st-order predicates of all types (for example, "is red" and "="), and second, he extended it to 1st-order quantifiers (2nd-order predicates) of all types (for example, "is a relation between humans" and "is symmetric").

Lindström's criterion for logical quantifiers and predicates can be formulated as follows:

(ISO) A Quantifier/predicate Q is logical iff for any Q-structure S, Q is invariant under all isomorphisms of S.

By a Q-structure S we mean a structure $<A, \beta_1, \ldots, \beta_n>$, where A is a nonempty set (the universe of S), and $\beta_1, \ldots, \beta_n$ are elements or structures of elements of A of the same types as the corresponding arguments of Q. Let $S = <A, \beta_1, \ldots, \beta_n>$ and $S' = <A', \beta'_1, \ldots, \beta'_n>$. We say that structures S and S' are isomorphic—$S \cong S'$—iff there is a bijection (a 1–1 and onto function) f from A to A' such that for $1 \leq i \leq n$, $f^*(\beta_i) = \beta'_i$, where f^* is induced by f. Q is invariant under all isomorphisms of Q-structures S iff for any Q-structures S, S', if $S \cong S'$, then $Q_A(\beta_1, \ldots, \beta_n) = Q_{A'}(\beta'_1, \ldots, \beta'_n)$.

Among the new quantifiers/predicates satisfying ISO are, in addition to all the above examples of quantifiers satisfying PER, the identity predicate of standard 1st-order logic as well as nonstandard quantifiers such as the 2-place Most and Only, and the 1-place Symmetric and Well-Ordered. As the last two examples suggest, logical quantifiers, according to ISO, are not limited to (although they include all) cardinality quantifiers. Examples of a predicate and a quantifier which do not satisfy ISO are "is red" and "is a relation between humans".

We can indicate the type of a given quantifier by an n-tuple representing the types of its arguments, using <1> for a property on, or a subset of, the universe; <1,1> for a pair of subsets of the universe; <2> for a 2-place relation on the universe, and so on. Quantifiers of types involving only 1's we call "predicative", those of types involving at least one number larger than 1 we call "relational" or "polyadic". Thus, the standard $\forall$ and $\exists$ are predicative quantifiers of type <1>, but the 2-place Most is a predicative quantifier of type <1,1>, and the 1-place relational (polyadic) Symmetric is of type <2>. The last two are defined, semantically, as follows:

Given a universe A, subsets B,C of A, and a binary relation R on A,
($M^{1,1}$) $(M^{1,1})_A(B,C) = T$ iff $|B \cap C| > |B - C|$.
(S^2) $(S^2)_A(R) = T$ iff R is symmetric.

They appear in English sentences such as:

(10) Most presidents are women.
(11) Equality is a symmetric relation.

Syntactically, these sentences have the logical forms:

(12) $(M^{1,1}x)(Px,Wx)$.
(13) $(S^{2}xy)\ Exy$.

Literally, (13) say, "Symmetric xy, x = y". It means something like "For x and y to be equal is for them to be symmetrically equal."

Tarski (1966) formulated essentially the same criterion, albeit in terms of permutations. As demonstrated by McGee (1996), a formulation in terms of isomorphisms is preferable.

Note: the presentation of this section is based on Sher (1991).

2 Logical Results

Generalized quantifier theory has had a considerable impact on the development of mathematical logic since the 1960s. It has led to the development of a new field, model-theoretic logic, and the establishment of numerous results. (For a representative collection of articles in this field see Barwise & Feferman 1985.) Two results of special interest are due to Lindström (1969, 1974) and Keisler (1970).

Lindström's Theorem

By offering a model-theoretic generalization of the notion of logical constant, Mostowski and Lindström created tools for comparing the strength of different 1st-order logical systems (with respect to their ability to define model-theoretic structures). These tools enabled Lindström to arrive at an important characterization of standard 1st-order logic:

(FOL) Standard 1st-order logic is the strongest logic that has both the compactness/completeness and the Löwenheim-Skolem properties (see Section II.1 above).

Keisler's Result

Keisler showed that completeness is not limited to standard 1st-order logic; some stronger logics are also complete. In particular 1st-order logic with the generalized quantifier Uncountably Many has a complete proof system. (In contrast, 1st-order logic with the quantifier At Least Denumerably Many does not.) Keisler's result weakens the case for the standard 1st-order quantifiers based on completeness (see II.1. and II.3. above).

3 Linguistic Results

Generalized quantifier theory has led to a renaissance in linguistic semantics. We can divide its linguistic contributions into two: contributions to the study of natural-language *determiners*, and contributions to the study of natural-language *polyadic (relational)*

quantifiers. In both cases the restriction to *logical* quantifiers is relaxed, but here we will limit ourselves to logical quantifiers.

(a) Determiners

Determiners form an important linguistic category of noun modifiers (different from adjectives). Determiners include *quantity* modifiers that are naturally construed as quantifiers. Some of these—for example, "all", "every", "each", "any", "a", "some", "at least (exactly, at most) one"—are naturally construed as the standard quantifiers, $\forall$ and $\exists$, or as quantifiers defined in terms of these. However, many quantitative determiners cannot be analyzed using the standard quantifiers. A few examples are "most", "few", "half", "more . . . than—", "finitely many", "uncountably many", "$\aleph_0$", and so on. Generalized quantifier theory enables linguists to account for the semantics of all these determiners by treating them as predicative quantifiers. Among the pioneers in this field are Barwise & Cooper (1981), Higginbotham & May (1981), van Benthem (1983), Westerståhl (1985), and Keenan & Stavi (1986).

Quantifiers representing determiners are for the most part of type <1,1>, and sometimes of more complex types, such as <<1,1>,1>. We have seen above how $\text{Most}^{1,1}$ is defined. $\exists^{1,1}$ is defined as follows:

($\exists^{1,1}$) $\exists^{1,1}_A(B,C) = T$ iff $|B \cap C| > 0$

We may say that $\exists^{1,1}$ is generated from $\exists^1$ by relativizing it to a subset of the universe, B. Most existential quantifications in English have this (or some derivative) form. In the simplest case, linguists assign

(14) Some B's are C's

the logical form

(15) $(\exists^{1,1}x)(Bx, Cx)$

rather than

(16) $(\exists^1x)(Bx \,\&\, Cx)$.

A determiner-quantifier of type <<1,1>1> appears in

(17) More students than teachers attended the lecture,

whose logical form is:

(18) $(O^{<1,1>,1}x)(Sx, Tx, Lx)$.

$O^{<1,1>,1}$ is defined by:

(19) $O^{<1,1>,1}_A(B,C,D) = T$ iff $|B \cap D| > |C \cap D|$.

Linguists and logicians have developed a rich theory of the characteristic features of determiner-quantifiers. For an up-to-date account see Peters & Westerståhl (2006), Chs. 3–6.

(b) *Relational or Polyadic Quantifiers*

Generalized quantifier theory enables us to go beyond determiners to more complex quantificational structures. The study of relational or polyadic quantifiers in natural language has been pursued by Higginbotham & May (1981), Keenan (1987, 1992), van Benthem (1989), Sher (1991), and many others. The simplest relational quantifiers are of type <2>. We have already defined one relational quantifier of this type, Symmetric.

One way to arrive at relational quantifiers is by combining predicative quantifiers in some order. In the simplest case, relational quantifiers of this kind correspond to *linear* quantifier-prefixes (of other quantifiers); in more complex cases they correspond to non-linear, *branching* or *partially-ordered* quantifier-prefixes.

An example of a relational quantifier equivalent to a linear prefix of predicative quantifiers is $\forall\exists^2$, where

(20) $(\forall\exists^2 xy)\ \Phi xy$

is logically equivalent to

(21) $(\forall^1 x)(\exists^1 y)\ \Phi xy$.

While $\forall\exists^2$ simply says "for every' x there is a y such that", a more intricate relational quantifier that resembles it but cannot be expressed by a linear prefix of ∀'s and ∃'s, says "Every x has its own y". This quantifiers appears, for example, in

(22) For every drop of rain that falls, a flower grows,
(23) Every family has its own misery. (Every family is unhappy in its own way. Tolstoy)

If we symbolize this quantifier by "$\forall\rightarrow\exists^2$", the logical structure of (22) and (23) is:

(24) $(\forall\rightarrow\exists^2 xy)(Bx, Cy)$.

Other interesting relational quantifiers appear in:

(25) Different students answered different questions on the exam. (Keenan 1987)
(26) No three students answered the same number of questions. (ibid.)
(27) Most pairs of people are not married to each other.

An up-to-date discussion of relational (polyadic) quantifiers appears in Peters & Westerståhl (2006), Ch. 10.

A special type of relational quantifier corresponds to a *branching* quantifier-prefix. But the branching prefix is more interesting when thought of as a complex prefix. Two examples of irreducibly branching quantifier-prefixes are:

$$\begin{matrix}(\forall x)(\exists y)\\ \\ (\forall z)(\exists w)\end{matrix}\ \Phi xyzw. \quad \text{(Henkin 1959)} \tag{28}$$

$$\begin{matrix}(M^1x)\\ \\ (M^1y)\end{matrix}\ \Phi xy. \quad \text{(Barwise 1979)} \tag{29}$$

English sentences exemplifying these forms, or in the case of (29), the more complex form:

$$\begin{matrix}(M^{1,1}x)\psi_1x\\ \\ (M^{1,1}y)\psi_2y\end{matrix}\ \Phi xy \quad \text{(Barwise 1979)} \tag{30}$$

are:

(31) Some relative of each villager and some relative of each townsman hate each other. (Hintikka 1973)

(32) Most boys in my class and most girls in your class have all dated each other. (Variation on Barwise 1979)

The branching form raises interesting questions of interpretation. In particular, there is no consensus on the interpretation of non-monotone-increasing branching quantifiers (Q^1 is monotone increasing iff $Q^1(B) = T$, $B \subseteq C \Rightarrow Q^1(C) = T$) and on how to formulate a general semantic definition of branching quantifications that will deal with branching prefixes of any complexity. Most importantly, the branching form raises the question of *compositionality*, i.e., the question of whether the interpretation of logically complex formulas is always determined by that of their simpler components. The key issue is the structure of the branching prefix, which need not be a tree structure (but only a partial ordering). In addition to the works mentioned above, see Westerståhl (1987) and Sher (1990, 1997). For a further development (IF languages) and a discussion of compositionality see Hintikka & Sandu (1989, 2001).

4 Philosophical Results

Generalized quantifier theory provides solutions to all the philosophical problems plaguing the standard theory that were discussed in Section II.3 above. How successful these solutions are, is an open question. But generalized quantifier theory undoubtedly makes an important contribution to philosophy by creating a framework for confronting philosophical questions that were earlier thought to resist any systematic investigation. This is especially true of the last three questions, (d)–(f), which concern fundamental issues in the philosophy of logic.

Briefly, the solutions to the six problems raised in Section II.3 are:

(a) Basic (Non-standard) Logical Inferences

The problem of recognizing and accounting for basic inferences based on quantifier-notions is largely, and perhaps fully, solved by the generalized theory. All inferences

based on such quantitative notions as "most", "few", "half", "finitely many", "as many as", etc., can be accounted by this logic.

(b) Logical Expressive Power

This problem is also largely (perhaps fully) solved by generalized quantifier theory. *All* common quantitative notions, including all quantitative determiner notions, can be construed as logical notions in this theory. Indeed, every classical mathematical notion can be construed as a logical quantifier, predicate, or function according to this theory, often by transforming it to a 2nd-order notion. As a result, the expressive power of generalized 1st-order logic as a whole is very large. (See McGee 1966.)

(c) Formal Expressive Power & the Löenheim-Skolem Theorem

In many systems of generalized 1st-order logic the Löenheim-Skolem theorem does not hold. As a result, formal or mathematical notions that lose their intended meaning in standard 1st-order logic retain their intended meaning in such systems. There are two possible ways of achieving this result: by constructing such notions as logical, and by constructing them as nonlogical within an appropriate generalized logical system. The first was noted in (b) above. We can, for example, construct the formal notion "uncountably many" as a logical quantifier, UNC^1, defined, for an arbitrary universe A by:

(UNC^1) $UNC^1_A(B) = T$ iff $|B| > \aleph_0$.

Since the interpretation of logical constants is built into the apparatus of models (see Sher 1991), UNC^1 will preserve its intended meaning in every model. Sometimes, however, we prefer to define a formal notion as a nonlogical notion (for example, in order to explain its meaning or establish its properties). In this case we might be able to construct an appropriate generalized framework that blocks the Löenheim-Skolem theorem, and define the desired concept within that framework without subjecting it to the Löenheim-Skolem effect. (The gain in expressive power is not totally free; often, the price includes loss of compactness and other properties that make standard 1st-order logic easier to work with.)

(d) A Theoretical Criterion for Logical Quantifiers

Generalized quantifier theory provides a criterion (necessary and sufficient condition) for logical quantifiers and predicates—ISO—that is as theoretical, systematic, precise, and comprehensive as is the truth-functional, Boolean criterion for logical connectives. Both criteria are essentially invariance criteria. The truth-functional criterion says that a sentential connective is logical iff it does not distinguish anything about its arguments besides their bare truth value. ISO says that a quantifier/predicate is logical iff it does not distinguish anything about its arguments besides their bare formal structure. That is, logical connectives are invariant under all truth-preserving variations in their arguments; logical quantifiers are invariant under all structure-preserving variations in their arguments. Both criteria are also precise mathematical criteria: The sentential criterion is based on Boolean algebra, the quantificational criterion is based on set theory

(most commonly ZFC). Finally, both criteria determine a maximal collection of logical constants, connectives, and quantifiers, respectively. In so doing, both sanction the existence of a far larger number of logical constants than one would anticipate based on (mere) intuition, indeed an infinite number of logical connectives and quantifiers. (The fact that we cannot reduce the collection of logical quantifiers to a small finite collection as in the case of the logical connectives is to be expected. The power of quantificational logic is so much greater than that of sentential logic that it cannot have as simple a logical apparatus as the latter's.)

(e) Philosophical Understanding of Logicality

For reasons indicated in the discussion of this topic in Section II.3 above, a theoretical criterion for logical constants can be instrumental in providing a philosophically informative characterization of logicality. What is the philosophical content of ISO? The idea of invariance under isomorphisms is a systematic rendering of the intuitive idea of being indifferent to, or not distinguishing, the individual characteristics of objects in any given universe of discourse. There is a very long tradition of viewing this idea as characterizing logic. Two examples, taken from Kant and Frege, are:

> [General logic] treats of understanding without any regard to difference in the objects to which the understanding may be directed.
>
> (Kant 1981/7: A52/B76)

> Pure logic . . . disregard[s] the particular characteristics of objects.
>
> (Frege 1879: 5)

ISO offers a precise characterization of this condition and places it at the center of our understanding of logic. To be logical is to abstract from, or disregard, everything about objects and their relations besides their formal features or structure. That is, to be logical is to be formal or structural in a very strong sense. (It is, in a certain sense, to be maximally formal or structural.)

This characterization explains some of the intuitive features of logic: its high degree of necessity, extreme generality, topic neutrality, strong normativity, and even apriority (or quasi-apriority). Very briefly, since logical constants are formal, the laws governing them are necessary. Since logical constants apply regardless of what objects exist in the domain (universe of discourse), their laws are highly general and topic neutral. Since they are not affected by the empirical features of objects (or the laws governing these features), their laws are (largely) nonempirical. And since their invariance is greater than that of physical and other constants (which are not invariant under *all* isomorphisms), the laws associated with them are normatively stronger than those associated with other constants (logic does not have to obey the laws of physics, but physics has to obey the laws of logic). For sources and further discussion see Tarski (1966), Sher (1991, 1996, 2008), McGee (1996), and others.

(f) Foundation of Logical Consequence

By providing a general criterion for logical constants (combining ISO for predicates and quantifiers and the Boolean criterion for logical connectives), generalized

quantifier theory enables us to evaluate the adequacy of Tarski's semantic definition of logical consequence. The question is whether Tarski's model-theoretic test yields intuitively necessary and formal consequence in languages with logical constants satisfying the above criterion. The answer appears to be positive. Since (i) Tarskian consequences depend on the logical constants of the given language and on the apparatus of models; since (ii) according to the above criterion, logical constants are formal (in the sense of taking into account only the formal or structural features of their arguments); and since (iii) Tarskian models represent all formally possible situations relative to the given language, consequences based on formal constants satisfying Tarski's test are necessary and formal.

For a view claiming that the problem of logical constants is a "red herring", see Etchemendy (1990). For a discussion of this and the other issues raised in (d)–(f), see Sher (1991, 1996, 2008).

5 *Criticisms and Alternatives*

To date, generalized quantifier theory has been criticized mostly on two grounds.

Some of the criticisms (for example, Hanson 1997 and Gómez-Torrente 2002) concern the applicability of ISO to natural language quantifiers. The critics present examples of natural-language quantifier expressions that purportedly satisfy ISO yet are intuitively nonlogical. They conclude that the characterization of logical constants should be pragmatic rather than theoretical.

Other criticisms (for example, Feferman 1999) focus on more theoretical issues, like the relation between logic and mathematics and certain meta-mathematical features of quantifiers satisfying ISO (although here too, "gut feeling" plays a central role). Two logicians (Feferman 1999 and Bonnay 2008) offer alternatives to ISO, for example, invariance under homomorphisms (Feferman). For response to some of these criticisms see Sher (2001, 2003, 2008).

Related Topics

1.3 Logical Form
1.8 Compositionality
3.4 Quantifiers and Determiners
3.6 Anaphora
4.4 The Role of Mathematical Methods
5.1 Model Theory
5.9 Montague Grammar
7.4 Frege, Russell, and Wittgenstein

References

Barwise, J. (1979) "On Branching Quantifiers in English," *Journal of Philosophical Logic* 8: 47–80.

Barwise J. & R. Cooper (1981) "Generalized Quantifiers and Natural Language," *Linguistics & Philosophy* 4: 159–219.

Barwise, J. & S. Feferman (eds.) (1985) *Model-Theoretic Logics*, New York: Springer-Verlag.

Bonnay, D. (2008) "Logicality & Invariance," *Bulletin of Symbolic Logic* 14: 29–68.

Dummett, M. (1973) "The Philosophical Significance of Intuitionistic Logic," in *Truth and Other Enigma*, Cambridge, MA: Harvard University Press, 1978, pp. 215–47.

Etchemendy, J. (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press.
Feferman, S. (1999) "Logic, Logics, and Logicism," *Notre-Dame Journal of Formal Logic* 40: 31–54.
Frege, G. (1879) "Begriffsschrift, a Formula Language, Modeled upon That of Arithmetic, for Pure Thought," in J.v. Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Cambridge, MA: Harvard University Press, 1967, pp. 1–82.
Gärdenfors, P (ed.) (1987) *Generalized Quantifiers: Linguistic & Logical Approaches*, Dordrecht: D. Reidel.
Geach, P.T. (1962) *Reference & Generality: An Examination of Some Medieval & Modern Theories*, Ithaca: Cornell University Press.
Gentzen, G. (1935) "Investigations into Logical Deduction," in M.E. Szabo (ed.) *The Collected Papers of Gerhard Gentzen*, Amsterdam: North-Holland, 1969, pp. 68–131.
Gödel, K. 1930. "The Completeness of the Axioms of the Functional Calculus of Logic", in *Collected Works*, Vol. I, Feferman et al. (eds.), New York: Oxford University Press, 1986, pp. 103–123.
Gómez-Torrente, M. (2002) "The Problem of Logical Constants," *The Bulletin of Symbolic Logic* 8: 1–37.
Hanson, W.H. (1997) "The Concept of Logical Consequence," *Philosophical Review* 106: 365–409.
Henkin, L. (1959) "Some Remarks on Infinitely Long Formulas", in *Finitistic Methods: Proceedings of the Symposium on Foundations of Mathematics*, Warsaw: P.W.N. and Pergamon Press, 1961, pp. 167–83.
Higginbotham, J. & R. May (1981) "Questions, Quantifiers and Crossing," *Linguistic Review* 1: 41–79.
Hintikka, J. (1973) "Quantifiers vs. Quantification Theory," *Dialectica* 27: 329–58.
Hintikka, J. & G. Sandu (1989) "Informational Independence as a Semantic Phenomenon," in J.E. Fenstad et al. (eds.), *Logic, Methodology, and Philosophy of Science*, Amsterdam: Elsevier, pp. 571–89.
—— (2001) "Aspects of Compositionality," *Journal of Logic, Language, and Information* 10: 49–61.
Kamp, H. (1981) "A Theory of Truth & Semantic Representation,"in J. Groenendijk et al. (eds.), *Truth, Interpretation, and Information*, Dordrecht: Foris, 1984, pp. 1–41.
Kant, I. (1981/7) *Critique of Pure Reason*, Tr. N. Kemp Smith, London: Macmillan, 1929.
Keenan, E.L. (1987) "Unreducible *n*-ary Quantifiers in Natural Language," in Gärdenfors 1987: 109–50.
—— (1992) "Beyond the Frege Boundary," *Linguistics and Philosophy* 15: 199–221.
Keenan, E.L. & J. Stavi (1986) "A Semantic Characterization of Natural Language Determiners," *Linguistics & Philosophy* 9: 253–329.
Keisler, H.J. (1970) "Logic with the Quantifier 'There Exist Uncountably Many'," *Annals of Mathematical Logic* 1: 1–93.
Lindström, P. (1966) "First Order Predicate Logic with Generalized Quantifiers," *Theoria* 32: 186–95.
—— (1969) "On Extensions of Elementary Logic," *Theoria* 35: 1–11.
—— (1974) "On Characterizing Elementary Logic," in S. Stenlund (ed.), *Logical Theory and Semantic Analysis*, Dordrecht: D. Reidel, pp. 129–46.
Marcus, R.B. (1972) "Quantification and Ontology," *Noûs* 6: 240–50.
May, R. (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press.
McGee, V. (1996) "Logical Operations," *Journal of Philosophical Logic* 25: 567–80.
Montague, R. (1973) "The Proper Treatment of Quantification in Ordinary English," in *Formal Philosophy: Selected Papers*, R.H. Thomason (ed.), New Haven: Yale University Press.
Mostowski, A. (1957) "On a Generalization of Quantifiers," *Fundamenta Mathematicae* 44: 12–36.
Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.
Partee, B.H. (1987) "Noun Phrase Interpretation and Type-Shifting Principles," in J. Groenendijk et al. (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Dordrecht: Foris, pp. 115–143.
Peters, S. & D. Westerståhl (2006) *Quantifiers in Languag eand Logic*, Oxford: Oxford University Press.
Priest, G. & R. Routley (1989) "The Philosophical Significance and Inevitability of Paraconsistency," in G. Priest et al. (eds.) *Paraconsistent Logic Essays on the Inconsistent*, Munich: Philosophia Verlag, pp. 483–539.
Quine, W.V. (1970) *Philosophy of Logic*, 2nd edition, Cambridge, MA: Harvard University Press 1986.
Rescher, N. (1962) "Plurality-Quantification," Abstract, *Journal of Symbolic Logic* 27: 373–4.
Sher, G. (1990) "Ways of Branching Quantifiers," *Linguistics and Philosophy* 13: 393–422.
—— (1991) *The Bounds of Logic: A Generalized Viewpoint*, Cambridge, MA: MIT Press.
—— (1996) "Did Tarski Commit 'Tarski's Fallacy'?," *Journal of Symbolic* Logic 61: 653–86.
—— (1997) "Partially-Ordered (Branching) Generalized Quantifiers: A General Definition," *Journal of Philosophical Logic* 26: 1–43.
—— (2001) "The Formal-Structural View of Logical Consequence," *Philosophical Review* 110: 241–61.
—— (2003) "A Characterization of Logical Constants *Is* Possible," *Theoria* 18: 189–97.

—— (2008) "Tarski's Thesis," in D. Patterson (ed.), *New Essays on Tarski and Philosophy*, Oxford: Oxford University Press, pp. 300–39.
Tarski, A. (1933) "The Concept of Truth in Formalized Languages". In J. Corcoran (ed.) *Logic, Semantics, Metamathematics*, 2nd ed. Indianapolis: Hackett, 1983, pp. 152–278.
—— (1966) "What Are Logical Notions?" *History and Philosophy of Logic 7* (1986): 143–54.
Van Benthem, J. (1983) "Determiners and Logic," *Linguistics and Philosophy* 6: 447–78.
—— (1989) "Polyadic Quantifiers", *Linguistic sand Philosophy* 12: 437–64.
Westerståhl, D. (1985) "Logical Constants in Quantifier Languages," *Linguistics and Philosophy* 8: 387–413.
—— (1987) "Branching Generalized Quantifiers and Natural Languages," in Gärdenfors, pp. 269–98.

5.3
THE LOGIC OF TIME AND TENSE

Antony Galton

1 Introduction

Classical logic makes no special provision for handling the expression of time and temporal phenomena; yet time undeniably plays an important part in language, surfacing both in the grammar (as tense and aspect) and in the lexicon (notably in the form of adverbs such as "now", "then", "always", and prepositions and connectives such as "before", "after", "since", "until"). For logic to be of service to the analysis of language, ways must be found of expressing the formal content of such linguistic items within a logical framework.

We may distinguish two broadly different approaches to increasing the expressive range of a logical system; the difference between them relates to the distinction between the logical and nonlogical elements of a logical language. In standard first-order predicate logic, the *logical vocabulary* comprises the Boolean connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$, and $\rightarrow$, the quantifiers $\forall$ and $\exists$, and, for first-order logic with identity, the symbol $=$. These symbols have fixed interpretations: that is, their meanings are determined by the definition of the logic, not by the user. By contrast, the interpretation of individual constants, function symbols, and predicate symbols is not determined in the logic itself; rather, the user is free to specify their interpretations, subject only to the constraint that they adhere to the correct form laid down by the logic: An individual constant must be interpreted as denoting an element of the domain of discourse $\triangle$, an n-ary function symbol as denoting a function of type $\triangle^n \rightarrow \triangle$, and an n-ary predicate symbol as denoting a set of n-tuples of domain elements. Thus the formula $\forall x(P(x) \rightarrow \exists y(Q(y) \wedge R(x, y)))$ might be understood as saying that every positive integer has a prime factor, that every human being has a mother, or that every country has a capital city (among innumerable other possibilities), depending on how the nonlogical symbols P, Q, and R are interpreted; but throughout all the possible interpretations, the logical symbols $\forall$, $\exists$, $\wedge$, and $\rightarrow$ are interpreted in exactly the same way.

If we wish to use a logic to handle some new range of expressive possibilities, then there are, in general, two approaches we may adopt:

1. To adhere to the original logic but capture the new range of meanings by extending the interpretations of the nonlogical vocabulary.
2. To modify the logic itself by introducing new logical constants whose fixed meanings relate to the desired additional expressive power.

The latter approach is appropriate if the new meanings to be covered are sufficiently universal in their applicability, featuring in a wide range of specific subject domains. This accords with the traditional view of logic as *transcendental*, that is, transcending any specific subject matter. The first approach, by contrast, suggests itself when the new meanings clearly belong to some specific domain of discourse: If, for example, we wished to investigate the logic of kinship relations, it would be natural to introduce predicates or functions whose interpretations correspond to natural language terms such as "mother", "brother", or "cousin", without seeking to modify the inherent expressive and inferential apparatus of the formal logical system employed.

In the case of time and temporal phenomena, the choice between the two approaches is finely balanced. On the one hand, temporality is pervasive, permeating much of our language, reflecting the "universal" character of time as the medium in which events and processes occur. On the other hand, much of the historical development of logic, especially in the last two centuries, is inseparable from its use as a medium for formalizing mathematics, a domain of discourse from which the temporal element is conspicuously lacking. At least from a Platonistic perspective, mathematical truths are eternal and unchanging: our knowledge and understanding of those truths may change, but for a Platonist these are in no way constitutive of the truths themselves. From this point of view, time and temporality may seem to lack the degree of universality required for a topic to be enshrined in the fixed logical vocabulary of a logic.

As a result of this finely-balanced choice, the historical development of temporal logic has followed two different approaches, each with its own adherents, and a considerable body of work devoted to exploring their relative merits and the relations between them. Under the first approach, the expression of temporality is accomplished through the use of a first-order language whose nonlogical vocabulary includes elements which are to be given a temporal interpretation; this approach encompasses a range of rather different possibilities, which are discussed in section 2. The second, more radical, approach goes beyond first-order logic by introducing modal-type operators as the primary vehicle for the expression of temporality. This approach originated with *tense logic*, which is described in section 3.

2 First-order Approaches to the Logic of Time

The task here is to express the temporal content of propositions using a logic which does not contain resources specifically designed for that purpose—in particular, standard first-order predicate calculus. The most obvious way of doing this, the *method of temporal arguments* (MTA), is described in the first subsection below; subsequent subsections cover alternatives to MTA, in which content normally thought of as propositional in nature is *reified*.

2.1 The Method of Temporal Arguments

In the method of temporal arguments, any atomic proposition whose truth value can vary with time is equipped with an additional argument-place to hold a reference to the time at which, in any particular occurrence, it is to be evaluated. To represent the sentence "It is raining," for example, instead of a simple 0-ary predicate Raining, we could use a 1-ary predicate Raining(–), where the argument place is to be filled by a term denoting a time, as e.g., Raining(t_1), where t_1 refers to, say, 11.45 p.m. (GMT) on March 30 2010.

Using this style of analysis, one can express a wide range of temporal propositions, for example:

"It is raining whenever we are in Scotland"	$\forall t(\mathsf{In}(\mathsf{we}, \mathsf{scotland}, t) \rightarrow \mathsf{Raining}(t))$
"John was in London all last week"	$\forall t(\mathsf{During}(t, \mathsf{lastWeek}) \rightarrow \mathsf{In}(\mathsf{john}, \mathsf{london}, t))$
"John kissed Mary last Tuesday"	$\exists t(\mathsf{During}(t, \mathsf{lastTuesday}) \wedge \mathsf{Kiss}(\mathsf{john}, \mathsf{mary}, t))$
"We are going to Scotland in July"	$\exists t(\mathsf{During}(t, \mathsf{july}) \wedge \mathsf{GoTo}(\mathsf{we}, \mathsf{scotland}, t))$
"John has never kissed Mary"	$\neg\exists t(\mathsf{Before}(t, \mathsf{now}) \wedge \mathsf{Kiss}(\mathsf{john}, \mathsf{mary}, t))$

These are "naive" renderings, with no attempt at capturing delicate semantic nuances, and the reader will doubtless find cause to question them: The point, however, is to give an initial indication of the general style of analysis to be expected from MTA. Some of the nuances are discussed below.

The chief elements of this kind of representation are:

1. Explicit references to specific times, represented by individual constants (lastWeek, lastTuesday, july, now). As presented, these appear to be indexical (since, e.g., the interval referred to by "last week" depends on when it is uttered), but in principle they could be replaced by non-indexical expressions such as "the week beginning April 12, 2010"—although it is a question of some philosophical moment whether temporal discourse can be completely purged of indexicality without loss of content.
2. Generalized temporal reference via the quantified variable t.
3. Relations between times, as expressed by the predicates During and Before.
4. "Thematic" content, describing specific states of affairs or events, e..g., Kiss(john, mary, –).

Formally, we have a first-order language whose nonlogical vocabulary comprises sets T of individual constants denoting times, R of temporal relations, and P of monadic predicates over times. The predicates in P may have additional nontemporal structure (e.g., Kiss(john, mary, –) is further decomposed into a ternary predicate Kiss(–,–,–) and individual constants john and mary). Reasoning with such a language is supported by axioms specifying properties of time. For example, the temporal precedence relation might be stipulated to be transitive:

$$\forall t\forall t'\forall t''(\mathsf{Before}(t, t') \wedge \mathsf{Before}(t', t'') \rightarrow \mathsf{Before}(t, t''))$$

and irreflexive:

$$\forall t\neg\mathsf{Before}(t, t).$$

Should these individual times be instants or intervals? Technically, instants are easier to handle. If the flow of time is linear, distinct instants t and t' must stand in one of two relations: either t precedes t' or t' precedes t. By contrast, two distinct intervals can stand

in any of *twelve* different relations, depending on how the endpoints of one are related to the endpoints of the other. Since an interval is determined by the two instants marking its endpoints, any representation that uses intervals can, in principle, be converted into one which only uses instants. For conformity with everyday language, however, this may not be the most sensible choice. A common expression of temporality is the location of an event within an interval: "I went to London yesterday", "His father died in 1998", and so on. Any formalism lacking direct representation of intervals must resort to paraphrase to express such facts. Indeed, it has been suggested that the notion of an instant does not correspond to any empirically observable phenomenon, and that *all* references to individual times should be to intervals, which derive their empirical reality from being the times at which events happen (Allen 1984).

How is the thematic content of a temporal proposition related to the times to which it is ascribed? If t is an instant, then In(john, london, t) says that John is in London at that instant. Since being in London necessarily persists over an interval, that instant must form part of some interval throughout which John is in London. If, on the other hand, t is an interval, then In(john, london, t) could be understood as saying either that John is in London *throughout* t, or that he is there *at some time* during t. Ordinary language does not come down decisively one way or the other. John's being in London is a state; further complications arise when we consider events, as in Kiss(john, mary, t). Compare "John kissed Mary yesterday" and "John kissed Mary at midnight". In the former, the kiss takes place at a time falling within the interval designated *yesterday*, without being coextensive with it. In the latter, the kiss occupies a—presumably short—interval which contains, or starts at, midnight.

To use the formal apparatus of MTA to represent the meanings of ordinary language sentences, we must be sensitive to issues such as these. The flexibility of natural language makes this enterprise difficult: there is seldom any tight one-to-one correspondence between the meanings of natural language sentences and the meanings expressible within a precisely defined logical system. Hence it may not be appropriate to try to express all linguistic meanings in logical form; the virtue of logical analysis is in providing a means for expressing precisely, within a mathematically watertight framework, a range of possible meanings, in order to provide a set of controlled exemplars with reference to which the vagaries of natural language can be discussed.

2.2 Reified Temporal Logics I: Type-reification

For the purposes of reasoning about temporal information, one drawback of MTA is that it does not allow one to express in first-order form general statements such as "An effect cannot precede its cause" or "A state of affairs that obtains throughout the day must obtain throughout the morning", which play a part, at least implicitly, in everyday temporal reasoning. The need to accommodate such statements has been keenly felt in artificial intelligence (AI), where the task of automating reasoning forces one to tease out many general assumptions underlying ordinary reasoning but not explicitly expressed.

A widely-advocated solution within AI is the use of a "reified" notation for expressing temporal facts. Where MTA writes, e.g., Kiss(john, mary, t), the reified system of (Allen 1984), has Occurs(kiss(john, mary), t). Here the thematic content expressed in MTA by the predicate Kiss(john, mary, _) is expressed by means of a *term* kiss(john, mary) referring to an event-type, here conceived as an entity in its own right (hence "reified"),

to which is ascribed the property of occurring at various times. To express a causal relationship between two events, Allen writes Ecause(e, t, e', t'), meaning that an occurrence of event-type e at time t caused an occurrence of event-type e' at t'. If t and t' are instants, then the rule that an effect cannot precede its cause could be expressed by

$$\forall e \forall e' \forall t \forall t' (\text{Ecause}(e, t, e', t') \rightarrow \neg \text{Before}(t', t)).$$

The reified approach allows explicit distinctions between states, processes, and events. For Allen, an event forms a unitary whole, no proper part of which can be an event of the same type. This is consistent with the categories of "accomplishment" and "achievement" in the well-known classification of (Vendler 1967). In reified logic, this may be expressed by the axiom:

$$\forall e \forall t \forall t' (\text{Event}(e) \wedge \text{Occurs}(e, t) \wedge \text{ProperPart}(t', t) \rightarrow \neg \text{Occurs}(e, t')).$$

For this to make sense, Occurs(e, t) must mean that an occurrence of event-type e is exactly coextensive in time with the interval t. This in turn means that our example "John kissed Mary last Tuesday" must be expressed as

$$\exists t (\text{ProperPart}(t, \text{lastTuesday}) \wedge \text{Occurs}(\text{kiss}(\text{john}, \text{mary}), t)).$$

This could be abbreviated to

$$\text{OccursIn}(\text{kiss}(\text{john}, \text{mary}), \text{lastTuesday}),$$

as advocated in (Galton 1990).

In contrast to events, states do not form discrete individual occurrences, instead exhibiting the property of *dissectivity*:

$$\forall s \forall t \forall t' (\text{State}(s) \wedge \text{Holds}(s, t) \wedge \text{ProperPart}(t', t) \rightarrow \text{Holds}(s, t')).$$

For example, if it was raining all day yesterday, then it was raining at lunchtime yesterday. Note the use of "Holds" rather than "Occurs": for Allen, the ways in which states and events are realized at times are sufficiently distinct to warrant the use of different predicates to express them.

2.3 Reified Temporal Logics II: Token-reification

Although type-reification provides useful expressive power, it may be objected to on philosophical grounds, since it seems to accord event and state types first-class status as elements of one's ontology. An alternative form of reification, which countenances individual event occurrences but not general event types, is sometimes proposed instead. This is known as *token-reification*.

Token-reification can be traced back to Davidson's well-known paper on the logical form of "action sentences" (Davidson 1967). This addressed the problem of the apparent "variable polyadicity" of verbs. To give a simple example, the English verb "sing" can be used intransitively, as in "John sang", or transitively, as in "John sang *Rule Britannia*".

Naïvely, one might try representing these by Sing(john) and Sing(john, ruleBritannia). In logic, however, a predicate has a fixed number of argument places (its *polyadicity*), so the two occurrences of Sing here must represent different predicates—Sing_1 (john) and Sing_2 (john, ruleBritannia), say. The problem then arises of accounting for the logical relation between the two predicates: specifically, that Sing_2 (john, ruleBritannia) implies Sing_1 (john).

In this case there is a simple solution: Any singing must be a singing of something (even if only a random sequence of notes), so Sing_1 (x) could be *defined* to mean $\exists y \mathsf{Sing}_2$ (x, y). But now consider a sequence such as

John sang.
John sang *Rule Britannia*.
John sang *Rule Britannia* in the bathroom.
John sang *Rule Britannia* in the bathroom at midnight.

where each sentence is implied by the ones following it. One might suggest using a four-place predicate Sing_4, representing the sentences by the formulae

$\exists x \exists y \exists z \mathsf{Sing}_4$(john, x, y, z)
$\exists y \exists z \mathsf{Sing}_4$ (john, ruleBritannia, y, z)
$\exists z \mathsf{Sing}_4$ (john, ruleBritannia, bathroom, z)
Sing_4 (john, ruleBritannia, bathroom, midnight)

but this is unsatisfactory since there may be no limit to the number of extra arguments needed (e.g., consider "John sang *Rule Britannia* loudly in the key of $E\flat$ in the bathroom at midnight"), making it impossible to fix the polyadicity of predicates in advance as logic requires.

Davidson's suggestion was to treat these sentences as referring to a particular individual event, to which are ascribed various properties: that it was a singing by John of *Rule Britannia*, its location was the bathroom, its time was midnight, it was loud, its key was $E\flat$, etc. Using an existentially quantified variable "e" to denote this event, we can then express the last example using the formula

$\exists e(\mathsf{Sing}(\mathsf{john}, \mathsf{ruleBritannia}, e) \wedge \mathsf{Location}(e, \mathsf{bathroom}) \wedge \mathsf{Time}(e, \mathsf{midnight}) \wedge \mathsf{Loud}(e) \wedge \mathsf{Key}(e, \mathsf{eFlat}))$.

Formulae corresponding to the shorter sentences above then follow as logical consequences from this by standard first-order logic. Since "e" here represents an individual occurrence, or event-token, this is an example of token-reification as defined earlier.

In computer science, a similar method was proposed by Kowalski & Sergot (1986), who introduced the "event calculus" as a system for reasoning about events in a logic-based framework, with intended application to database updates and discourse representation. Vila & Reichgelt (1996) also advocated a form of token-reification for temporal reasoning in AI.

To round off the discussion of first-order approaches to time, consider how the sentence "Napoleon invaded Russia in 1812" might be represented in the three approaches:

- Method of Temporal Arguments: Invade(napoleon, russia, 1812).

- Type-reification: OccursIn(invade(napoleon, russia), 1812).
- Token-reification: ∃e(Invade(napoleon, russia, e) ∧ OccursIn(e, 1812)).

For further discussion and comparison of these methods and their variants, see (Galton 2006).

3 Modal Approaches to the Logic of Time

From a linguistic point of view, one drawback of the first-order approaches is that the relationship between the expression of propositions in natural language and their logical representations can be very indirect. The sentence "When John arrived, Mary had already left", for example, might be represented using MTA as

∃t∃t'(Arrive(john, t) ∧ Leave(mary, t') ∧ Past(t) ∧ Before(t', t)),

or using token-reification as

∃e∃e' ∃t∃t'(Arrive(john, e) ∧ Leave(mary, e') ∧ Time(e, t) ∧ Time(e, e') ∧ Past(t) ∧ Before(t', t)).

To offer either of these formulae as an analysis of the English sentence is to claim that it refers, implicitly, to unspecified times t and t', and, in the reified analysis, to individual events e and e'. Thus these analyses impute to the English sentence a significant ontological load. Conversely, some elements expressed directly in the English sentence, notably the simple past tense in "arrived", the pluperfect tense in "had left", and the temporal relation indicated by "when", receive no direct expression in the logic, their senses being distributed across different elements of the representation. For some purposes, a logical representation which more directly mirrors the forms of ordinary language, such as tense, might be preferred. This is one of the motivations underlying the modal approach to time.

3.1 Tense Logic

Tense logic was the creation of Arthur Prior and arose directly from modal logic via a proposal to reduce modal notions to temporal ones, defining the possible as "what is or will be", and the necessary as "what is and always will be", which Prior attributed to the third-century BCE logician Diodorus Cronus. If modality is defined in this way, what existing system of modal logic is obtained? To determine this, we need a logic of futurity, and this led directly to the construction of a "calculus of tenses" (Prior 1967).

In tense logic, the standard logical apparatus is supplemented by *tense operators*, which are applied to propositional formulae to produce new formulae. Prior's original logic has four operators, with intended meanings as follows:

Pϕ: It has been the case that ϕ
Hϕ: It has always been the case that ϕ
Fϕ: It will be the case that ϕ
Gϕ: It will always be the case that ϕ

where ϕ stands for any proposition. The operators **P** and **F** correspond roughly to the simple past and future tenses of natural language: if ϕ stands for "It is raining", $\mathbf{P}\phi$ would stand for "It was raining", and $\mathbf{F}\phi$ for "It will be raining". Because these operators form propositions out of propositions, they can be iterated, leading to "compound tenses" such as $\mathbf{PF}\phi$ ("It was going to rain"), $\mathbf{FP}\phi$ ("It will have rained"), and $\mathbf{PP}\phi$ ("It had rained"). Using this formalism, our example "When John arrived, Mary had already left" might be represented as $\mathbf{P}(\mathrm{Arrive(john)} \wedge \mathbf{P}\mathrm{Leave(mary)})$, where the simple past tense of "John arrived" corresponds to the first **P** operator, while the pluperfect tense of "Mary had left" corresponds to the nesting of the second **P** operator within the scope of the first.

The relationship between the "weak" **P** and **F** operators and their "strong" counterparts **H** and **G** is exactly analogous to that between the weak and strong modal operators $\Diamond$ and $\Box$. Thus $\mathbf{P}\phi$ may be defined as $\neg\mathbf{H}\neg\phi$, since to say that it has been true that ϕ is precisely to deny that it has always been false that ϕ, and likewise $\mathbf{F}\phi$ is equivalent to $\neg\mathbf{G}\neg\phi$. Conversely, $\mathbf{G}\phi$ and $\mathbf{H}\phi$ are equivalent to $\neg\mathbf{P}\neg\phi$ and $\neg\mathbf{F}\neg\phi$ respectively. Thus in principle we can dispense with **P** and **F** and express everything in terms of **H** and **G**, or vice versa.

Prior's approach to formalizing the logic of tenses was *axiomatic*: the meanings of the operators were not given explicitly but implicitly through the use of *axioms*, that is, tense-logical formulae postulated as necessarily true. A minimal set of axioms, making no presuppositions about the structure of time, is:

1. $\phi \rightarrow \mathbf{GP}\,\phi$
2. $\phi \rightarrow \mathbf{HF}\,\phi$
3. $\mathbf{H}(\phi \rightarrow \psi) \rightarrow (\mathbf{H}\phi \rightarrow \mathbf{H}\psi)$
4. $\mathbf{G}(\phi \rightarrow \psi) \rightarrow (\mathbf{G}\phi \rightarrow \mathbf{G}\psi)$

Axiom 1 asserts that whatever is true now will always have been true (the past is irrevocable), and axiom 2 asserts that whatever is true now has always been going to be true. The latter may smack of fatalism, but arguably, in accepting it, we are not committing ourselves to asserting that present facts were inevitable, merely that, as things turned out, they became true. Philosophically, this is already highly problematic, bearing as it does on the problem of future contingents, which has exercized philosophers at least since Aristotle. The remaining axioms are less contentious: Axiom 3 states that if ϕ has always implied ψ, then if ϕ has always been true, so has ψ; and Axiom 4 the analogous statement for the future.

To use these axioms in reasoning, one requires also some *rules of inference*. In addition to a standard set of rules for classical logic, one requires rules (RH and RG) which say that if ϕ can be derived from the axioms, then so can $\mathbf{H}\phi$ and $\mathbf{G}\phi$: in other words, having proved some formula, one can assert that it always has been and will be true. The system thereby obtained is called *minimal tense logic*, designated K_t. An example of a theorem of K_t is

$$\phi \wedge \mathbf{P}\psi \rightarrow \mathbf{P}(\psi \wedge \mathbf{F}\phi),$$

e.g., if it is raining and was sunny, then it has been the case that it was sunny and would be raining. (For a proof of this formula, see (Galton 1984, p.43).)

Many plausible tense-logical formulae cannot be derived as theorems of K_t, and it is

natural to ask which additional formulae should be postulated as axioms. The formula $\mathbf{FF}\phi \rightarrow \mathbf{F}\phi$, for example, says that whatever will be future is already future. This seems sensible since what will count as future from the perspective of some future time must surely form part of what counts as future now. Although this seems obvious, it makes a substantive assumption about the organization of time: that the relation "is in the future of" is transitive. The converse formula, $\mathbf{F}\phi \rightarrow \mathbf{FF}\phi$, says that anything currently in the future will still be in the future at some future time. This also makes a substantive assumption about time. For suppose time t is immediately followed by t', with no other times between them; if ϕ is true at t' but at no other times, then at t, $\mathbf{F}\phi$ is true but $\mathbf{FF}\phi$ is false, falsifying $\mathbf{F}\phi \rightarrow \mathbf{FF}\phi$. Hence to posit this formula as an axiom is, implicitly, to rule out a model of time in which one time can immediately follow another—and thus to assume a *dense* model, in which between any distinct times there is another (and hence, by iteration, infinitely many). Both these formulae, then, make substantive assumptions about time, and as such go beyond K_t.

Quite early in the development of tense logic, serious expressive limitations became apparent. For example, Prior's tense operators cannot be used to express sentences such as "Mary will be unhappy until John arrives" or "John has been unhappy since Mary left." For this reason, Kamp (1968) introduced two new binary operators **S** and **U**, whose intended meanings are as follows:

> $\mathbf{S}\,\phi\psi$: It has been the case that ϕ, and between then and now it has always been the case that ψ.
> $\mathbf{U}\phi\psi$: It will be the case that ϕ, and between now and then it will always be the case that ψ

Informally, these can be read as "Since ϕ, ψ" and "Until ϕ, ψ", and the two sentences above can be expressed as **U**Arrive(john)Unhappy(mary) and **S**Leave(mary)Unhappy(john). Note, however, that whereas the former formula implies FArrive(john), the English sentence does not imply that John will arrive; a more satisfactory rendering of that sentence might be **U**Arrive(john)Unhappy(mary) $\vee$ **G**Unhappy(mary). Having introduced **S** and **U**, we can dispense with Prior's operators, since $\mathbf{P}\phi$ and $\mathbf{F}\phi$ can be defined as $\mathbf{S}\phi(\phi \vee \neg\phi)$ and $\mathbf{U}\phi(\phi \vee \neg\phi)$ respectively, with **G** and **H** defined in terms of **P** and **F** in the usual way.

4 Relationships Between the Approaches

4.1 Model Theory for Tense Logic

Although the two approaches to the logic of time may appear very different, there are important links between them. In presenting the operators used in the modal approach, reference was made to their "intended meanings", without any indication of how operator and meaning are linked. In general terms, how are meanings assigned to logical formulae? Under the axiomatic approach, this is done implicitly, by specifying the logical relationships between the formulae (i.e., what can be deduced from what, and what is true unconditionally). The alternative is to assign meanings to formulae explicitly by specifying how they are to be interpreted, as was done informally above. This may be made formal through the discipline of *formal semantics*, typically in the form of a *model theory* mapping elements of the logic to abstract structures which serve as models for

real-world situations represented by the formulae.

We earlier observed that the formula $\mathbf{FF}\phi \rightarrow \mathbf{F}\phi$, if asserted as an axiom, implicitly presupposes that the relation "is in the future of" is transitive. To make such statements explicit, we interpret formulae with respect to a *temporal frame* $\mathrm{T}_{<}$ comprising a set T of individual times (assumed to be instants) ordered by a *precedence* relation $\prec$: thus $t \prec t'$ means that time t precedes time t'. (Note that we use "t" to refer to a time in a temporal frame, as distinct from "$\mathbf{t}$" used earlier as a temporal variable in a first-order language.) Temporal frames are directly analogous to the frames used in modal logic. To evaluate formulae as true or false at different times, one must specify, for each atomic formula ϕ, its truth value $\mathrm{V}(\phi,\mathrm{t})$ at each time t in the temporal frame under consideration. This is called a *valuation* over the frame $\mathrm{T}_{<}$. An *interpretation* $I=(\mathrm{T}_{<},\mathrm{V})$ of the tense-logical language consists of a temporal frame together with a valuation. The meaning of a tense operator is specified by stating the conditions under which a formula containing that operator is satisfied (i.e., evaluated as true) at a time by any given interpretation. For the two future-tense operators, this is done as follows (and analogously for the past):

- I satisfies $\mathbf{F}\phi$ at t if and only if, for some $t' \in T$, $t \prec t'$ and I satisfies ϕ at t'.
- I satisfies $\mathbf{G}\phi$ at t if and only if, for every $t' \in T$, if $t \prec t'$ then I satisfies ϕ at t'.

We may also speak of a temporal *frame* satisfying a formula, meaning that every interpretation over that frame satisfies the formula at every time. Given these definitions, the theorems of minimal tense logic K_t are precisely those formulae satisfied in every temporal frame, thus justifying the claim that K_t makes no specific assumptions about the temporal ordering.

By contrast, the formula $\mathbf{FF}\phi \rightarrow \mathbf{F}\phi$, which is *not* a theorem of K_t, is satisfied by a temporal frame $\mathrm{T}_{<}$ if and only if its precedence relation is transitive, that is:

$$\forall t \forall t' \forall t''(t \prec t' \wedge t' \prec t'' \rightarrow t \prec t'').$$

To prove this, suppose first that $\prec$ *is* transitive, and suppose interpretation I satisfies $\mathbf{FF}\phi$ at t. This means there are times t', $t'' \in T$ such that $t \prec t'$, $t' \prec t''$, and I satisfies ϕ at t''. Since $\prec$ is transitive, $t' \prec t''$, so I satisfies $\mathbf{F}\phi$, and therefore also $\mathbf{FF}\phi \rightarrow \mathbf{F}\phi$, at t. Hence every interpretation over the frame satisfies this formula at all times, so the frame itself satisfies it. Conversely, suppose $\prec$ is *not* transitive: then there are times t', t'' such that $t \prec t'$, $t' \prec t''$, but $t \not\prec t''$. Define a valuation V over $T_{<}$ such that ϕ is true at t'' and at no other times. Then the interpretation $(T_{<},V)$ satisfies $\mathbf{FF}\phi$ at t (since it satisfies ϕ at a time in the future of a time in the future of t), but it does not satisfy $\mathbf{F}\phi$ at t (since the only time ϕ is satisfied is not in the future of t). Hence it does not satisfy $\mathbf{FF}\phi \rightarrow \mathbf{F}\phi$ at t, so the frame does not satisfy this formula either.

There are many such cases where a tense-logical formula ϕ is satisfied by a temporal frame $T_{<}$ if and only if the frame satisfies some specified first-order condition. Another example is the formula $\mathbf{F}\phi \rightarrow \mathbf{FF}\phi$, which is satisfied by *dense* temporal frames, satisfying the first-order condition

$$\forall t \forall t'(t \prec t' \rightarrow \exists t''(t \prec t'' \wedge t'' \prec t')).$$

The systematic investigation of these correspondences between tense-logical (and

more generally modal) formulae and the first-order conditions on frames which satisfy them is called *correspondence theory*; for details, see (Van Benthem 1991). Not all tense-logical formulae correspond to first-order conditions in this way, and not all first-order conditions on temporal frames correspond to tense-logical formulae. No tense-logical formula, for example, is valid on all and only the *irreflexive* frames, which satisfy the condition $\forall t \neg(t \prec t)$.

Using the model theory, we can systematically translate tense-logical formulae into first-order formulae, in the style of either MTA or a reified system. Thus **PRaining** (i.e., "It has been raining") translates as $\exists$t(Before(t, now) $\wedge$ Raining(t)) or $\exists$t(Before(t, now) $\wedge$ Holds(raining, t)).

4.2 Hybrid Tense Logic

The correspondences described in the preceding section suggest one way in which the two approaches to the logic of time might be unified: the first-order approach provides a way of specifying the semantics of formulae produced under the modal approach, enabling tense-logical formulae to be translated into first-order formulae. This does not in itself overcome certain apparent expressive limitations of the modal approach, and in this section we examine how a closer *rapprochement* between the two approaches might be achieved, enabling at least some first-order formulae to be expressed in modal form.

A deficiency of tense logic, in either Prior's original form or Kamp's extended version, is that it does not lend itself readily to expressing sentences containing explicit reference to times, such as "John kissed Mary at midnight". One possibility is to treat "midnight" not as an individual term, but as a statement, expressing the proposition "It is midnight." Then the formula P(Midnight $\wedge$ Kiss(john, mary)) states that at some past time it was midnight and John kissed Mary. If Midnight is understood in a generic sense as true whenever it is midnight, this does not locate the kiss at a specific midnight occasion; but if instead we introduce a proposition $\text{Midnight}_{20100412}$, which is only true at midnight on April 12 ,2010, then we can write P($\text{Midnight}_{20100412}$ $\wedge$ Kiss(john, mary)) to say that John kissed Mary at that time.

This idea lies at the heart of *hybrid tense logic* (HTL), which marries the spirit of tense logic with the explicit temporal references of the first-order approach. In addition to "ordinary" propositions, HTL includes special propositions called *nominals* which can only be evaluated as true at exactly one time. A nominal such as $\text{Midnight}_{20100412}$ may be thought of as "naming" the unique time at which it is true. In what follows we use t, t′ t" etc. for nominals, and ϕ, ψ, etc. for general propositions.

To assert that ϕ holds at the time designated by nominal t, with no assumption as to whether t is past, present, or future, one can put $\mathbf{P}(\mathsf{t} \wedge \phi) \vee (\mathsf{t} \wedge \phi) \vee \mathbf{F}(\mathsf{t} \wedge \phi)$, which is abbreviated to $@_{\mathsf{t}}\,\phi$. We can regard $@_{\mathsf{t}}$ as a kind of tense operator, although unlike the usual tense operators it commutes with negation, i.e., $\neg @_{\mathsf{t}}\phi$ is equivalent to $@_{\mathsf{t}} \neg\phi$. To specify the semantics of @ in the model-theoretic style, it is convenient to write g(t) to denote the unique time at which t is true. Then we can put

- *I* satisfies $@_{\mathsf{t}}\phi$ at t if and only if *I* satisfies ϕ at *g*(t).

Superficially, $@_{\mathsf{t}}\phi$ looks like a syntactic variant of a type-reified formula of the form Holds(s, t) (see section 2.2). The parallel is not exact, however, since whereas ϕ is a

propositional formula, s is a term denoting a state. Hence whereas operators of the form @t can be iterated to yield formulae such as @t @t' ϕ, Holds cannot: Holds(Holds(s, t'), t) is ill-formed. This difference need not be serious, however, since from the semantics of @, the formula @t @t' ϕ is equivalent to @t' ϕ, so all but the last of a string of @ operators can be dropped without change of meaning.

The introduction of nominals changes the complexion of tense logics considerably. For example, in ordinary tense logic the formula $\mathbf{F}(\phi \wedge \psi) \wedge \mathbf{F}(\phi \wedge \chi) \rightarrow \mathbf{F}(\psi \wedge \chi)$ is not valid, but if ϕ is an HTL nominal then it is. Again, in the model theory, the class of all irreflexive frames is exactly characterized by the hybrid formula $\mathsf{t} \rightarrow \neg \mathsf{F}\mathsf{t}$, but not by any ordinary tense-logical formula.

There is no space here to discuss HTL further. It is the temporal form of hybrid logic, which is related to modal logic just as HTL is related to tense logic. For further details, see (Areces, Blackburn & Marx 2001).

5 Concluding Remarks

In the space available it has not been possible to do more than outline the main ways formal logic has been used for representing time and tense. The logic of time is of interest to researchers in a number of different areas, notably philosophy, linguistics, computer science, and mathematics, and therefore it has been approached from many different points of view. See the references and further reading for some pointers.

Related Topics

2.8 Event Semantics
5.4. Modal Logic.

References

Allen, J. (1984), 'Towards a general theory of action and time,' *Artificial Intelligence* 23, 123–54.

Areces, C., Blackburn, P. & Marx, M. (2001), 'Hybrid logics: Characterization, interpolation and complexity', *Journal of Symbolic Logic* 66(3), 977–1010.

Davidson, D. (1967), The logical form of action sentences, *in* N. Rescher, ed., *The Logic of Decision and Action*, University of Pittsburgh Press, Pittsburgh, Reprinted in D. Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

Galton, A. P. (1984), *The Logic of Aspect: an Axiomatic Approach*, Clarendon Press, Oxford.

Galton, A. P. (1990), 'A critical examination of Allen's theory of action and time,' *Artificial Intelligence* 42, 159–88.

Galton, A. P. (2006), 'Operators vs arguments: The ins and outs of reification', *Synthese* 150, 415–441.

Kamp, J. A. W. (1968), Tense Logic and the Theory of Linear Order, PhD thesis, University of California, Los Angeles.

Kowalski, R. A. & Sergot, M. J. (1986), 'A logic-based calculus of events,' *New Generation Computing* 4, 67–95.

Prior, A. N. (1967), *Past, Present, and Future*, Clarendon Press, Oxford.

Van Benthem, J. F. A. K. (1991), *The Logic of Time: A Model-Theoretic Investigation into the Varieties of Temporal Ontology and Temporal Discourse*, 2nd. edn, Kluwer Academic Publishers, Dordrecht.

Vendler, Z. (1967), Verbs and times, *in Linguistics and Philosophy*, Cornell University Press, Ithaca, chapter 4. Reprinted in Mani *et al.* (2005).

Vila, L. & Reichgelt, H. (1996), 'The token reification approach to temporal reasoning', *Artificial Intelligence* 83, 59–74.

Further Reading

Fisher, M., Gabbay, D. & Vila, L., eds, (2005), *Handbook of Temporal Reasoning in Artificial Intelligence*, Elsevier, Amsterdam. (An emphasis on formal treatments directed towards computational implementation.)

Mani, I., Pustejovsky, J. & Gaizauskas, R., eds, (2005), *The Language of Time: A Reader*, Oxford University Press, Oxford. (A useful compendium of articles covering different approaches to handling time in natural language.)

Øhrstrøm, P. & Hasle, P. F. V. (1995), *Temporal Logic: From Ancient Ideas to Artificial Intelligence*, Kluwer Academic Publishers, Dordrecht. (Offers a useful historical perspective.)

5.4
MODAL LOGIC AND ITS APPLICATION TO THE PHILOSOPHY OF LANGUAGE

Kit Fine

Modal logic is the logic of possibility and necessity and of other such notions. It began, as did logic in general, with Aristotle, in his theory of the 'modal syllogism'; and various notions and principles of modal logic were extensively discussed in the middle ages. But the subject only came into its own at the beginning of the twentieth century (see Goldblatt 2005 for an account of its recent history).

I begin by presenting some basic material on the possible worlds' approach to modal logic and then show how it relates to certain key topics in the philosophy of language. For reasons of space, I have had to be very selective and, inevitably, a great deal of interesting material has not been covered.

1 Language and logic

We begin by describing the language of sentential modal logic and the various systems that can can be formulated by its means.

1.1 Language

The language L of classical (non-modal) sentential logic may be taken to consist of the following symbols:

(i) the sentence letters p_1, p_2, p_3 . . .;
(ii) the truth-functional connectives $\sim$, $\vee$, $\wedge$ and $\supset$;
(iii) the parentheses (and).

1.2 Formulas

The formulas of L are generated by means of the following rules:

(i) each sentence letter is a formula;
(ii) if A is a formula then so is ~A;
(iii) if A and B are formulas, then so are $(A \vee B)$, $(A \vee B)$, $(A \wedge B)$ and $(A \supset B)$.

The language $L(\Box)$of modal sentential logic is obtained by adding a modal operator $\Box$ for 'necessity' to the language L. There is one additional rule for generating formulas:

(iv) If A is a formula, then so is $\Box A$.

We adopt the following abbreviations:

$\Diamond A$ ('possibly A') for $\sim \Box \sim A$;
$A \rightarrow B$ ('A strictly implies B') for $\Box(A \supset B)$.

We follow standard notation in using '$\supset$' for material implication and in using '$\rightarrow$' for various stricter forms of implication. It should be noted that embeddings of $\Box$ are allowed. Thus $\Box\Box A$ or $\Box A \supset \Box\Box A$ or $\Box(\Box A \supset A) \supset \Box A$ are all formulas. A large part of modal logic is concerned with the behavior of embedded modality.

1.3 Systems

A *modal system* is characterized by a set of axioms and rules. Say that a formula of $L(\Box)$ is *truth-functionally valid* if it is a substitution-instance of a tautology of L. The formula $(\Box A \vee \sim \Box A)$, for example, is truth-functionally valid since it is a substitution-instance of the tautology $(p_1 \vee \sim p_1)$. The axiom and rules for the *minimal* modal system K are then as follows (where A and B are any formulas of $L(\Box)$):

Truth-functionality.	All truth-functionally valid formulas;
Distribution.	$\Box(A \supset B) \supset (\Box A \supset \Box B)$;
Modus Ponens.	$A, A \supset B / B$;
Necessitation.	$A/\Box A$.

The *theorems of* the system K are all those formulas that can be obtained from the axioms by means of the rules. To be more exact, the set of K-theorems is the smallest set to contain the axioms of Truth-functionality and Distribution and to be closed under the rules of Modus Ponens and Necessitation (thus if A and A e B are in the set then so is B and if is A is in the set then so is $\Box A$). We may write '$\gamma \supset A$' to indicate that A is a theorem of K (and similarly for other systems of modal logic). What motivates the choice of K is that it is the weakest system to which the semantic techniques described below will apply.

2 Semantics

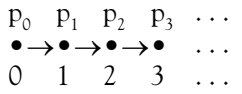
We now describe the possible worlds semantics for modal logic.

2.1 Models

A K-*model* M is an ordered triple (W, R, φ), where W is a non-empty set, R is a binary relation on W (i.e. $R \subseteq W \times W$), and φ is a function taking each sentence-letter p_1, p_2,

p_3 . . . into a subset of *W*. Intuitively, *W* is the set of 'worlds', relative to which formulas are evaluated, *R* is an accessibility-relation on worlds, holding between worlds *w* and *v* when *v* is relevant to the evaluation of a necessity formula at *w*, and φ is a 'valuation' that specifies, for each sentence-letter p, the set of worlds φ(p) at which p is true.

An example of a K-model is pictured below:



And it may be formally defined by letting:

W = {0, 1, 2, . . .};
R = {(m, n): n = m + 1}; and
φ = {(p_n, {n}): n = 0, 1, 2, . . .}.

Thus in this model, each sentence-letter p_n is true at exactly the one world n and the only world accessible from a given world m is its 'successor' n + 1. (We might think of the worlds as 'days', starting with day 0 and continuing into the infinite future; accessibility is the 'tomorrow' relation; and p_n is sentence stating that it is day n.)

2.2 Truth

Given a K-model M =(*W*, *R*, φ), we may define when a formula is true at a world of the model. We use the notation 'w ⊨ A' to indicate that A is true at w (or '*w* ⊨ $_M$A' if we wish to be explicit about the underlying model M). The definition is as follows:

M(i) *w* ⊨ p iff *w* ∈ φ(p), for any sentence-letter p;
M(ii) *w* ⊨ ~B iff it not the case that *w* ⊨ B;
M(iii) *w* ⊨ B ∨ C iff *w* ⊨ B or *w* ⊨ C (and similarly for ∧ and ⊃);
M(iv) *w* ⊨ □B iff *v* ⊨ B whenever *wRv*.

It is readily determined on the basis of these clauses and the abbreviations that have been laid down that:

M(v) *w* ⊨◊B iff *v* ⊨ B for some *v* for which *wRv*; and
M(vi) *w* ⊨B → C iff *v* ⊨ C whenever *wRv* and *v* ⊨ B.

A possibility formula ◊B is true at a world just in case the embedded formula B is true at some accessible world; and a strict implication B → C is true at a world just in case the consequent C is true in any accessible world in which the antecedent B is true.

Let us illustrate with the example-model above (in which we might think of '□' as meaning *tomorrow*). The sentence-letter p_n will be true at the world n (i.e. n ⊨ p_n) given that n ∈ φ(p_n). Since n–1*R*n for n > 0, it follows that ◊p_n is true at n-1 (i.e. n-1 ⊨◊p_n). Use ◊^{m} for ◊◊ . . . ◊ (m times). It then follows, for n > m, that ◊^{m} p_n is true at n–m (i.e. n–m ⊨ ◊^{m} p_n). The reader may readily establish the converse, namely that k ⊨ ◊^{m} p_n only if k = n–m; and so ◊^{m} p_n ⊃ ◊^{l} p_n will be false at the world n–m whenever l ≠m.

2.3 Validity

We say that a formula A is *true in* a model $M = (W, R, \varphi)$ if it is true in every world $w \in W$ of the model; and we say that a formula A is K-*valid* if it is true in every model. We write $\models_K A$ to indicate that the formula A is K-valid. We may show, for example, that any instance $\Box(A \supset B) \supset (\Box A \supset \Box B)$ of the Distribution axiom is K-valid. For take any model $M = (W, R, \varphi)$ and world $w \in W$ and suppose that $w \models \Box(A \supset B)$ and $w \models \Box A$. We must then show that $w \models \Box B$. But if $w \models \Box(A \supset B)$, then $v \models B$ whenever $v \models A$ and wRv; and if $w \models \Box A$, then $v \models A$ whenever wRv. So $v \models B$ whenever wRv; and consequently $w \models \Box B$, as required. On the other hand, the formula $\Diamond^m p_n \supset \Diamond^l p_n$, for $n > m$ and $l \neq m$, is not K-valid since it is not true at the world n–m in the example-model above.

3 Soundness and completeness of K

It is a central result concerning K that its theorems and validities coincide. It is useful to divide this result and its proof into two parts.

Theorem 1 (Soundness of K) Any K-theorem is K-valid.

Proof The proof is by induction on the generation of the K-theorems from the K-axioms by means of the K-rules. We must first show that the K-axioms are valid and that the K-rules preserve validity. We have already taken care of the Distribution axioms; and the other cases are equally straightforward.

Theorem 2 (Completeness for K) Any K-valid formula is a K-theorem.

The proof of this result is more difficult than for Soundness. Let us briefly sketch one method of proof, using a so-called 'Henkin' model.

Say that a set Δ of formulas is K-*consistent* if $\bot$ (defined as $\sim\top$) is not K-deducible from Δ; and say that a set Δ of formulas is *maximally* K-consistent if it is K-consistent but not properly contained in any K-consistent set. It is the maximally K-consistent sets that will become the worlds of the canonical model. The following result is familiar from the standard completeness results for classical sentential and quantificational logic:

Lemma 3 (Lindenbaum's Lemma) Every K-consistent set of formulas Δ is included in a maximal K-consistent set of formulas Γ.

For Δ set of formulas, let $\Box(\Delta) = \{A: \Box A \in \Delta\}$. We define the *canonical* K-*model* $M_K = (W_K, R_K, \varphi_K)$ *for* K by:

(i) $W_K = \{\Delta: \Delta$ is a maximally K-consistent set of formulas$\}$;
(ii) $R_K = \{(\Delta, \Gamma): \Delta, \Gamma \in W_K$ and $\Box(\Delta) \subseteq \Gamma\}$;
(iii) $\varphi_K = \{(p, V): p$ is a sentence-letter and $V = \{\Delta \in W_K: p \in \Delta\}\}$.

We think of the worlds in the model M_K as containing exactly those formulas that they make true. This motivates clause (iii), since we want p to be true in Δ just in case it is a member of Δ. It can also be seen to motivate clause (ii), for if $\Box A$ is true in the world Δ then we certainly want A to be true at any accessible world Γ and it is hard to see what further condition might reasonably be imposed on an accessible world.

It may be shown that the canonical model is faithful to our intentions:

Lemma 4 (Truth-lemma). Let Δ be any maximally K-consistent set. Then:
$A \in \Delta$ iff $\Delta \models A$ (in the model M_K).

Membership of a maximally consistent set coincides with truth in the canonical model.

Completeness for K (theorem 2) can now be established. For suppose that A is not a theorem of K. We may then find a maximally K-consistent set Γ containing $\sim A$. By the consistency of Γ, it follows that $A \notin \Gamma$ and so, by lemma 4, A is not true at world Γ in the canonical model M_K and hence A is not valid.

The proof of theorem 2 establishes two further results. Say that a set of formulas Δ is *satisfiable* if some model M and world w of the model, each formula A of Δ is true at the world in the model. Then:

Strong Completeness Every K-consistent set of formulas Δ is satisfiable.

Compactness Δ is satisfiable if each finite subset of Δ is satisfiable.

For given that Δ is K-consistent, we may obtain a maximally K-consistent set containing Δ and then proceed as before; and given that each finite subset of Δ is satisfiable, it will follow that Δ is K-consistent and hence satisfiable by Strong Completeness.

4 Some other systems

A similar result is readily established by similar means for various other systems of modal logic. Let us give an illustrative example.

4.1 Completeness for T

Let the system T be the result of adding all formulas $\Box A \supset A$ as axioms to K (each such axiom says that A is the case if A is necessarily the case). Say that a model $M = (W, R, \varphi)$ is a T-*model* if its accessibility relation R is reflexive, i.e. if wRw whenever $w \in W$; and say that a formula is T-*valid* if it is true in every T-model. We then have the following soundness result for T:

Theorem 5 (Soundness for T) Any T-theorem is T-valid.

Proof As for the soundness result for K. The one new case that needs to be considered is the T-validity of $\Box A \supset A$. So take any model $M = (W, R, \varphi)$ and any world $w \in W$, and suppose that $\Box A$ is true at w. Then A is true at v whenever wRv. We need to show that A is true at w. But R is reflexive; so wRw; and so A is true at w.

The proof of the results leading up to the completeness of K go through as before, but using T-relative notions in place of K-relative notions. This is because the proofs of these results depend only upon the *presence* of certain axioms and rules in K and not upon their *absence*. Thus they will go through for any other system that contains those axioms and rules. However, in order to establish the completeness of T, we also need to show:

Lemma 6 The canonical model $M_T = (W_T, R_T, \varphi_T)$ for T is a T-model.

Proof We need to show that R_T is reflexive, i.e. that $\Box(\Delta) \subseteq \Delta$ for any maximally T-consistent set of formulas Δ. So suppose that $\Delta \vdash_T \Box A$. Since $\Box A \supset A$ is a T-theorem, it follows that $\Delta \vdash_T A$; and so $A \in \Delta$, as required.

We therefore obtain:

Theorem 7 (Completeness for T) Any T-valid formula is a T-theorem.

And, of course, Strong Completeness and Compactness can also be obtained in the same way as before.

4.2 Other Extensions

A similar strategy can be employed for other systems. Suppose, for example, that we add the axiom:

4 $\Box A \supset \Box\Box A$

to K (or to T). It can then be shown that a formula will be a theorem of the resulting system K4 (or S4) just in case it is true at all worlds of all models whose accessibility relation is transitive (or reflexive and transitive).

A good number of more general completeness results can be established but the reader should be warned that not all modal logics are complete for some class of frames (see any standard text, such as Hughes & Cresswell (2009), for details).

5 Modal Predicate Logic

The addition of modality to first-order quantificational logic raises serious philosophical and technical problems. On the syntactic side, matters are relatively straightforward. To the rules for the generation of the formulas of first-order logic, we add clause 1.2(iv) above:

If A is a formula, then so is $\Box A$.

This allows $\Box\exists xPx$, $\exists x\Box Px$ and $\Box\exists x\Box Px$, for example, to be formulas.

5.1 The Problem of Quantifying In

There is a difficulty, first pointed out by Quine (1953), over the interpretation of formulas in which we 'quantify into' a modal context. Consider his example, 'for some number x it is necessary that x > 7' (of the form $\exists x\Box Px$). For this to be true, the open sentence 'necessarily x > 7' must be satisfied by some object. But whether this open sentence is satisfied by a given object depends upon how the object is described. Describe it as'9' and the open sentence is satisfied (since necessarily 9 > 7); describe it as 'the number of planets' and the open sentence is not satisfied (since it is not necessary that the number of planets > 7). Thus purely objectual satisfaction or quantification appear to make no sense in this case.

There is a corresponding problem about the satisfaction of a formula *at a world*. The formula 'x > 7' is satisfied by the number 9 in the actual world. But suppose I ask: is it

satisfied by the number in some other world? Whether it is would appear to depend upon how the number is to be identified in that world. Identify it through its position in the number series and the answer is 'yes'; identify it through its role in numbering the planets and the answer may be 'no.'

There have been three main semantical responses to this problem–the objectualist, the conceptualist, and the counterpart-theoretic. We consider each in turn and show how it leads to a distinctive style of semantics for quantified modal logic.

5.2 Objectualism

The objectualist disputes the data upon which the statement of the problem depends. He claims that it does make sense to say that the sentence 'necessarily x > 7' is satisfied by a particular number, independently of how it is described, and, similarly, he claims that there is no problem of identifying an object from one world to another; the object is simply given in each world as the object that it is. (Such an approach is championed in Kripke (1980).)

A simple objectualist semantics might go as follows. A *quantificational* K*-model* M is an ordered quadruple $M = (W, D, R, \varphi)$, with (W, R) a K-frame as before, D (domain) a non-empty set, and φ (valuation) a function taking each world $w \in W$ into a classical first-order model with domain D (thus φ_w will assign a subset of D^n to each n-ary predicate and similarly for any other non-logical constants of the language.

Given a model $M = (W, D, R, \varphi)$, a world $w \in W$ and an assignment α (taking each variable into an individual from D), we may now define what it is for a formula A to be true at w relative to the assignment α, which we symbolize as $w \vDash_\alpha A$:

(i) $w \vDash_\alpha Px_1 \ldots x_n$ iff $\langle \alpha(x_1), \ldots, \alpha(x_n) \rangle \in \varphi_w(P)$;
(ii) $w \vDash_\alpha \sim B$ iff not $w \vDash_\alpha B$, and similarly for the other truth-functional connectives;
(iii) $w \vDash_\alpha \forall xB$ iff $w \vDash_{\alpha'} B$ for any assignment α' that differs at most on x from α;
(iv) $w \vDash_\alpha \Box B$ iff $v \vDash_\alpha B$ whenever wRv.

The various forms of validity may then be defined in the usual way.

5.2 Conceptualism

The conceptualist grants that purely objectual quantification into modal contexts makes no sense but, rather than abandoning quantification into such contexts altogether, he replaces it by a corresponding form of quantification over individual concepts. Thus the quantifier 'for some number x' in 'for some number x, x > 7' is taken to range over number concepts, such as *the concept of being ninth in the number series* or *the concept of being the number of planets*; and the quantified statement will therefore be true since the open sentence 'necessarily, x > 7' is true when 'x' takes *the concept of being ninth in the number series* as its value. (Such an approach, in one of its forms, is championed by Carnap (1947).)

A *quantificational* K*-model* M is now an ordered quadruple $M = (W, D, C, R, \varphi)$, with (W, D, R, φ) as before and C (concepts) a non-empty set of functions from W into D. Thus each individual concept is identified with the function which takes each world into the individual that is picked out by the concept in that world. An assignment

α should now be taken to be a function from the variables into the members of C. Truth-relative-to-an-assignment ($w \models_\alpha A$) is then defined much as before, the principal difference being to the first clause:

(i)′ $w \models_\alpha Px_1 \ldots x_n$ iff $<\alpha(x_1)(w), \ldots, \alpha(x_n)(w)> \in \varphi_w(P)$.

In each world, the concepts behave in the manner of the individuals which they pick out.

A plausible condition on C is that, for every individual $i \in D$ and every world $w \in W$, there should be a concept c in C for which $c(w) = i$. The maximal set of concepts will be obtained when every function from W into D is required to belong to C. On the other hand, a minimal set of concepts may be obtained by subjecting C to the condition that $c(w) \neq d(w)$ whenever $c \neq d$; for each individual in each world there will be a 'privileged' concept which picks it out.

Partial concepts, i.e. partial functions from W into D, might also be allowed. The concept of *the sun*, for example, will not pick out any individual in a world without heavenly objects. The admission of partial concepts considerably complicates the semantics, since we need to decide how to modify (i)′ above when one of the arguments $\alpha(x_1)(w), \ldots, \alpha(x_n)(w)$ is not defined. Should $Px_1 \ldots x_n$ be taken to be false at w (relative to the assignment α) or 'gappy' (neither true nor false)? And if gappy, then how should the other clauses in the truth-*definition be modified* to deal with the gaps?

5.3 Counterpart Theory

Like conceptualism, counterpart theory rejects the intelligibility of objectual quantification into modal contexts. But rather than reconstructing quantification as quantification over concepts, it reconstructs it in terms of the tracking of individuals from one world to another. Suppose we are wondering whether a given individual satisfies 'it is necessary that x > 7'. Then according to the counterpart-theorist, we should determine whether each acceptable way of tracking the individual across different worlds preserves the truth of 'x > 7'. (Such an approach, in one of its forms is championed by Lewis (1968).).

A *model* M is now an ordered quadruple $M = (W, D, T, R, \varphi)$, where $(W, D, , R, \varphi)$ is as before and T (tracking) is a relation on (n+1)-tples of the form $(w, i_1, \ldots, i_n)$, with $n > 0$, $w \in W$ and $i_1, \ldots, i_n \in D$. T is subject to the conditions that it only relates (n+1)-tples to (n+1)-tples, that $(w, i_1, \ldots, i_n)T(v, j_1, \ldots, j_n)$ implies wRv, and that $(w, i_1, \ldots, i_n)T(v, j_1, \ldots, j_n)$ implies $(w, i_{p(1)}, \ldots, i_{p(n)})T(v, j_{p(1)}, \ldots, j_{p(n)})$ for any permutation p on {1, 2, . . ., n}. Intuitively, T relates $(w, i_1, \ldots, i_n)$ to $(v, j_1, \ldots, j_n)$ if one may acceptably reidentify $i_1, \ldots, i_n$ in w as $j_1, \ldots, j_n$ in v. Note that 'simultaneous' tracking is allowed; and one might also allow current tracking to depend upon past tracking, though with some complication to the semantics. A special case of tracking is by concepts. Thus, given a set of concepts C, we may take $(w, i_1, \ldots, i_n)T(v, j_1, \ldots, j_n)$ to hold if, for some concepts $c_1, \ldots, c_n$ in C, $c_1(w) = i_1, \ldots, c_n(w) = i_n$ and $c_1(v) = j_1, \ldots, c_n(v) = j_n$.

The truth-definition is as for the objectual semantics except that the clause for '$\Box$' requires appeal to tracking. Where the free variables of B are $x_1, \ldots, x_n$:

$w \models_\alpha \Box B$ iff $v \models_\beta B$ whenever wRv and $(w, \alpha(x_1), \ldots, \alpha(x_n))T(v, \beta(x_1), \ldots, \beta(x_n))$.

Thus as we go to the different accessible worlds v, we track the individuals $\alpha(x_1), \ldots, \alpha(x_n)$ that are in play in the given world w.

5.4 Variable Domains

In our presentation of the previous semantical schemes, it was presupposed that the domain D of quantification was constant from one world to another. But one may wish it to vary. If, for example, the quantifiers are to range over all actual objects or over all actual objects of a certain sort, then the domain should vary with what is actual.

A model under the objectualist approach (similar modifications may be made under the other approaches) will now be a quadruple (W, D, R, φ), with (W, R, φ) as before and D (domains) a function that takes each world $w \in W$ into a set D_w, at least one of which is non-empty. The clauses in the truth-definition are as before, but the clause for the quantifier takes the form:

> $w \vDash_\alpha \forall x B$ iff $w \vDash_\alpha B$ for any assignment α' that differs at most on x from α and for which $\alpha'(x) \in D_w$.

The quantifier ranges, in each world, over the *actual* objects of the world.

When the domain is constant, the formula $\forall x \Box B \equiv \Box \forall x B$ is valid, since the effect of either prefix, '$\forall x \Box$' or '$\Box \forall x$' is simultaneously to quantify over all worlds $w \in W$ and all individuals $i \in D$. But when the domain is allowed to vary, the validity of the formula may fail in either direction. Suppose first that wRv, that D_v is not included in D_w, and that P is a one-place predicate true in each world of the members of D_w. Then $\forall x \Box Px$ is true at w while $\Box \forall x Px$ is false at w. Suppose next that wRv, that D_w is not included in D_v and that P is a one-place predicate that is true in each world u of the members of D_u. Then $\Box \forall x Px$ is true at w while $\forall x \Box Px$ is false at w.

Suppose we have a variable-domain model (W, D, R, φ) as specified above. We may then interpret the quantifiers as ranging over all *possible objects* $D^+ = \bigcup_{w \in W} D_w$, i.e. over all objects that are actual in one or another possible world. The clause for quantified formulas is then:

> $w \vDash_\alpha \Pi x B$ iff $w \vDash_{\alpha'} B$ for any assignment α' that differs at most on x from α and for which $\alpha'(x) \in D_w$.

We might introduce an existence or actuality predicate E to single out the objects which are existent (or actual) in a given world:

> $w \vDash_\alpha Ex$ iff $\alpha(x) \in D_w$.

The actualist quantification $\forall x B$ may then be seen as a relativized form of possibilist quantification, $\Pi x(Ex \supset B)$.

5.6 Identity

We may add an identity predicate '=' to the language of quantified modal language. Under the objectual semantics, this will be governed by the following clause:

> $w \vDash_\alpha x = y$ iff $\alpha(x) = \alpha(y)$.

It should be evident that the formula $x = y \supset \Box(x = y)$ (the 'necessity of identity') will then be valid and that Ex might be defined by $\exists y(x = y)$.

Under the conceptualist semantics, we might take '=' to be governed by the clause:

$$w \models_{\alpha} x = y \text{ iff } \alpha(x) = \alpha(y)$$

requiring that the concepts be the same or by the clause:

$$w \models_{\alpha} x = y \text{ iff } \alpha(x)(w) = \alpha(y)(w)$$

requiring that the objects picked out by the concepts be the same. In the latter case, the formula $x = y \supset \Box(x = y)$ will no longer be valid since x and y may be assigned concepts (such as *the morning star* and *the evening star*) which only contingently pick out the same object.

The necessity of identity may also fail under the counterpart-theoretic semantics if it is allowed that the same object assigned to x and y can subsequently be tracked to different objects within a given possible world.

5.5 Completeness

The question of completeness is much more problematic for quantified modal logic than sentential modal, although many of the standard systems of quantified modal logic can be proved to be complete (see Brauner & Gilardi (2007) for a recent survey).

6 Applications to Specific Constructions

The modalities are open to a number of different interpretations. Let me briefly indicate what some of them are and how they may be of interest to the philosophy of language.

6.1 Alethic

Under the alethic interpretation, '$\Box$' is interpreted as necessary truth. However, different forms of necessity might be distinguished—the metaphysical, the natural, the logical, for example—and it is plausible that each of them should be subject to somewhat different axioms. Many philosophers have supposed that S5 (S4 plus the axiom $A \supset \Box \Diamond A$) is the correct logic for the notion of metaphysical necessity that has played such a prominent role in recent philosophical discussion; and in providing a semantics for this system, we may take the 'worlds' of a model to be metaphysical possible worlds and take the accessibility relation to be 'universal':

$w \models \Box A$ iff $v \models A$ for *every* world v in W.

6.2 Epistemic/Doxastic

There are two principal epistemic interpretations—'$\Box A$' (or 'KA') is read either as 'P knows that A' or as 'it is a priori that A'; while under the doxastic interpretation,

'□A' (or 'BA') is read as 'P believes that A' . We may take the worlds to be epistemically or doxastically possible worlds, worlds that are possible are far as it is known or believed; and we may take one world to be accessible from a given world if it is compatible with what is known or believed in the given world. It is often supposed that S4 is the correct logic of knowledge for the ideal knower and that K4 (possibly with the addition of □(□A ⊃ A)) is the correct logic for the ideal believer.

6.3 *Tense-logical*

There are two principal tense-logical interpretations—'□A' (or 'GA') is read as 'it will always be that A' and '□A' (or 'HA') is read as 'it always was that A'. In this case, the worlds should be taken to be instants of time and the accessibility-relation should be taken to be the earlier–later relation. Different views on the structure of the earlier–later relation correspond to different logics. When time is assumed to be linear, for example, the corresponding logic is K4.3 (K4 plus the axiom □(A & □A ⊃B) ∨□(A & □A ⊃B)) (Prior (1967)).

6.4 *Deontic*

Under this interpretation, '□A' (or 'OA') is read as 'it is obligatory that A' and its dual '◊A' (or 'PA') is read as 'it is permissible that A'. This interpretation is somewhat less natural than the others since obligation and permission are more naturally treated as a feature of acts rather than of propositions. Worlds are, perhaps, epistemically possible states and one world is accessible from a given world when it is compatible with what is obligatory in the given world. The system D (K plus the axiom □A ⊃◊A, signifying that the obligatory is permissible) is a natural choice of logic under this interpretation (Hilpinen (1970)).

6.5 *Vagueness-theoretic*

Suppose that Herbert is a borderline case of a bald man. Then we might express that this is so as the claim that it is not definite that he is bald and not definite that he is not bald. This suggests that we might interpret '□' as a definiteness operator D ('it is definitely the case that'). A world *w* may now be understood as a precification of the language, one under which each vague expression is made completely precise, and *R* may be understood as the relation that holds between *v* and *w* when *v* is an admissible precisification of the language relative to *w*. An interpretation of this sort has been proposed by Fine (1975) and similar interpretations have been pursued by Williamson (1994) and others. Different assumptions about the existence of 'higher order' vagueness will then lead to different logics for the operator D.

6.6 *Conditionals*

We have so far given different interpretations for the modalities '□' and '◊', but the possible worlds semantics can be applied to other kinds of operator as well. A prime example is that of the counterfactual conditional 'if A were to be the case, then C would be the case' (A > C). Given a formula A, let |A| be the set of worlds in which A is true (relative to a given model); and given a set of worlds *V* and a particular world

w, let $C(V, w)$ be the set of those worlds of V that are closest to w. Along the general lines of the semantics proposed by Stalnaker (1968) and Lewis (1973), we may then take:

$w \vDash A > C$ iff $v \vDash A$ for every world v in $C(|A|, w)$.

The counterfactual $A > C$ is to be true if its consequent is true in all of the closest worlds in which the antecedent is true.

The formal development of these various interpretations has been put to a large number of different uses, both inside and outside of philosophy. Of special interest to the philosophy of language is their role in understanding the corresponding locutions of ordinary language—of tense, modal auxiliaries, etc. But the applications have to made with a great deal of care since many aspects of ordinary use will be distorted or ignored. The sentence 'he does not know Obama is President', for example, embodies the presupposition that Obama is President though this is no part of the formal semantics for the knowledge operator or, again, 'Yesterday, John was ill' means that John's illness was yesterday not at some time prior to yesterday. But application of the symbolism and semantics of modal logic does provide an important initial guide to how the ordinary locutions are to be understood.

7 Application to More General Topics

Let me briefly discuss three more general areas of application within the philosophy of language.

7.1 Meaning

One naturally takes the aim of semantics to be the systematic assignment of a meaning to each meaningful expression of a given language. But what are meanings? The semantic value assigned to an expression by classical logic is merely its extension—the denotation in the case of a singular term, a set of individuals in the case of a one-place predicate, and a truthvalue in the case of a sentence. But expressions with the same extension may differ in their meaning. The meanings of 'the morning star' and 'the evening star', for example, are quite different even though their denotations are the same.

A more adequate counterpart to meaning may be obtained by letting the semantic value of an expression be its possible worlds 'intension', rather than its extension. The intension of an expression, intuitively speaking, is a function which takes each possible world into the object that would be the extension of the expression in that world. Thus the intension of a singular term will be an individual concept, a function taking each world into an individual, the intension of a one-place predicate will be a 'property', taking each world into a set of individuals, and the intension of a sentence will be a 'proposition', a function taking each world into a truth-value or what is, in effect, a set of possible worlds. Using intensions in place of extensions, we may discriminate more finely between the meanings of different expressions. The intensions of 'the morning star' and 'the evening star', for example, will not be the same since they will diverge on any world in which different 'stars' appear in the morning and in the evening.

A framework of this sort, in which extensions and intensions can be assigned to expressions of increasingly complex type, has been proposed as a general model for our understanding of ordinary language (Montague (1970)).

7.2 Rigid Intensions

Of special interest are the rigid or constant intensions. These pick out the same object in each possible world; and an expression may be said to be a rigid designator when its intension is rigid, i.e. when its designation is the same in each world. Each necessary truth, for example, will be a rigid designator since it will designate the Truth in each possible world.

The notion of rigid designation has become important in recent philosophy of language through work of Kripke (1980), in which he argued that proper names (such as 'Aristotle' or 'London') and certain common nouns (such as 'water', 'heat') were rigid designators while definite descriptions (such as 'the famous teacher of Alexander', 'the capital of London') were not.

7.3 Intensional Objects

Natural language appears to contain reference to intentional objects which do not appear to exist or, at least, do not exist in the usual way. It is correct to say, for example, that Newton was looking for the philosopher's stone, even though the philosopher's stone does not exist. This then raises the question of how such apparent reference to non-existents is to be handled. A natural suggestion within the possible framework is to treat such intentional objects as intensions. Thus the philosopher's stone would be a function taking each possible world into the object that is the philosopher's stone in that world (and perhaps into nothing if there is no such object).

7.4 Indexicality

The sentences of natural language are often context-dependent; their truth-value depends upon the context in which they uttered. So, for example, the truth-value of the sentence 'I am sitting' will vary with the speaker and the time of utterance.

The dependence of truth-value upon context is akin to the dependence of truth-value upon the world; and so we may use a general notion of index-relative evaluation to encompass these two forms of dependence. Thus a sentence such as 'I am sitting' may be taken to have a truth-value that is relative to a triple index (w, t, s), consisting of the world of evaluation w, the time of utterance t, and the speaker s. Different 'operators' may then pick up on these different indices. 'I might be sitting', for example, will be true at an index (w, t, s) if 'I am sitting' is true at an index (v, t, s) for some accessible world v; and 'I will be sitting' will be true at an index (w, t, s) if 'I am sitting' is true at (w, u, s) for some later time u.

A more interesting form of multiple indexing arises when we require two or more component indices of the same type. Consider 'you will regret what you are now doing'. In evaluating this sentence, 'will' will take us from the time of utterance t to a subsequent time u. But in evaluating the implicitly embedded 'you regret what you are now doing' at u, 'now' must be taken to refer, not to u, but to t. Thus the evaluation of the sentence requires that two temporal indices be kept in play—the 'fixed' time of utterance t and the 'floating' time of evaluation u.

7.5 *Content/Character*

The possible worlds framework has been helpful in distinguishing between different aspects of meaning. Suppose that I now say 'I am sitting' and that you now say 'I am sitting'. Then in one way, we say the same thing and, in another way, we do not. Use 'character' for the former and 'content' for the latter. The content of our utterances may then be represented by a set of possible worlds, the set of worlds in which I (in the one case) and you (in the other case) are now sitting, while the character of our utterances may be represented by the function which takes each context (as specified by a person and a time) into the corresponding content, i.e. that set of possible worlds in which the specified person is sitting at the specified time (Kaplan (1989)). And each of these is distinct again from the previous notion of content, as given by the set of triples (w, t, s) for which the speaker s is sitting at t in w.

7.6 *Limitations*

The possible worlds approach to meaning is subject to some well-known limitations. It cannot distinguish, for example, between knowing one necessary truth from knowing another. Or again, it may be permitted that I post the letter but not permitted that I post the letter or post the letter and burn down the post office, even though the two embedded clauses are true in the same possible worlds.

There is a question of how seriously to take these difficulties. My own view is that they cannot properly be overcome or ignored and that the possible worlds approach, for all of its success, can only be regarded as the first step towards a more adequate account of meaning.

References

Almog J., Perry J., Wettstein H. W. (1989) *Themes from Kaplan*, Oxford: Oxford University Press.

Blackburn P., Benthem J. van, & Wolter F. (eds.) (2007) *Handbook of Modal Logic*, Amsterdam: Elsevier.

Brauner T., Ghilardi S. (2007) 'First Order Modal Logic' in P. Blackburn et al. (2007), pp.549–620.

Carnap R. (1947) *Meaning and Necessity*, Chicago: University of Chicago Press.

Fine K. (1975) 'Vagueness, Truth and Logic', *Synthese* 30, 265–300; reprinted in *Vagueness: A Reader* (eds. R. Keefe and P. Smith), Cambridge: MIT Press (1996).

Goldblatt R. (2005) 'Mathematical Modal Logic: A View of its Evolution' in *Handbook of the History of Logic: VII* (eds. D. M. Gabbay and J. Woods), Amsterdam: Elsevier.

Hilpinen R. (ed.), (1970) *Deontic Logic: Introductory and Systematic Readings*', Dordrecht: D. Reidel.

Hughes G. F., Cresswell M. (2009) *A New Introduction to Modal Logic*, London: Routledge.

Kaplan D. (1989) 'Demonstratives' and 'Afterthoughts' in Almog et al. (1989).

Kripke S., (1980) *Naming and Necessity*, Cambridge: Harvard University Press.

Lewis D., (1973) *Counterfactuals*, Cambridge: Harvard University Press.

Lewis D. (1968) 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy*, 65 no. 5, 113–26.

Montague, R. (1960) 'On the Nature of Certain Philosophical Entities', *Monist*, 53, 159–194; reprinted in Thomason (1974), 148–187.

Montague, R. (1970). Pragmatics and Intensional Logic. *Synthèse*, 22, 68–94. Reprinted in Thomason (1974), 119–147.

Prior A. N. (1967) *Past, Present and Future*, Oxford: Oxford University Press.

Quine W. V. O. (1953) 'Reference and Modality' in *From A Logical Point of View*, Cambridge: Harvard University Press.

Stalnaker R. (1968) 'A Theory of Conditionals' in N. Rescher (ed.) *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series , No. 2', Oxford: Blackwell, 98–112.

Thomason, R. H. (ed.) (1974) *Formal Philosophy, Selected Papers of Richard Montague*. New Haven and London: Yale University Press.
Williamson T. (1994) *Vagueness*, London: Routledge.

Further Readings

Benthem. J. van (2010) *Modal Logic for Open Mind* Stanford: CSLI publications (overview, with applications to philosophy and computer science).
Chagrov, A., Zakyaryascher, M. (1997) *Modal Logic* Oxford: Clarendon Press (a more advanced text).
Fitting, M., Mendelsohn, R. (1998) *First-order Modal Logic*, Dordrecht: Kluwer (touches on many issues of interest to the philosopher of language).

5.5
TWO-DIMENSIONAL LOGIC AND TWO-DIMENSIONALISM IN PHILOSOPHY

Steven T. Kuhn

Introduction

The label "two-dimensional" has been applied to disparate logical systems. The first section of Humberstone 2004 nicely conveys the diversity. For present purposes, we may understand an n-dimensional logic to be one in which sentences acquire truth values relative to n parameters. Thus classical propositional logic, which represents sentences as being true or false *simpliciter*, is zero-dimensional. Common modal or tense logics, which represent sentences as being true or false relative to a world or a time, are one-dimensional. Logics aiming to represent both tense and modality assign truth values relative to both time and world. (Thomason 1984 provides an overview.) They are two-dimensional. For other linguistic phenomena, other parameters have been employed: for personal pronouns, a speaker, an addressee and a salience order among some class of individuals; for *here* and *there*, a pair of locations; for more faithful representations of tense (in a remarkable anticipation of ideas discussed here, dating to Reichenbach 1947), a time of utterance and a time referred to. It may sometimes be appropriate to view some parameters as constructed from others. To represent a notion of *historical* necessity according to which the past is determined and the future is open, worlds may be seen as constructed out of instants and their temporal order. Proper understanding of actions, events, states, and the like may require that we distinguish among sentences evaluated at instants and intervals and to view one as constructed from the other. For other applications it may be appropriate to view the parameters as primitive and perhaps independent. To represent tense and metaphysical necessity, we may take time and world as independent parameters. Alternatively, we may take time to be something internal to each world, allowing for the possibility that it has different structures in different worlds. All such generalizations of standard modal and tense logics may be regarded as many-dimensional, the dimension corresponding to the number of parameters to which the truth of a formula is relativized.

Recent interest in "two-dimensionalism" in philosophy stems from the observation that parameters of the kind mentioned above may play two distinct roles in the

determination of a sentence's truth value. The nature of these roles and the parameters that are supposed to play them is delineated differently by different authors, but it may be convenient to use the label 2D for two-dimensional logics whose parameters are both possible worlds. We thereby encompass the frameworks of Davies and Humberstone 1980 and Stalnaker (1981, 2004), an approximation to that of Chalmers (2006, 2006a, forthcoming), and illustrative fragments of those of Lewis 1981 and Kaplan (1978). We label the dimensions *utterance* world and *evaluation* world, or simply (to avoid prejudicing later interpretations) *u-world* and *e-world*. From a technical perspective, much of the interest in logics of this kind stems from the fact that the homogeneity of the two parameters allows each to take the value of the other. From a philosophical perspective, the dual role of possible worlds has been said to illuminate some of the most deep and central topics in contemporary philosophy: the existence and nature of contingent *apriori* and necessary empirical truths, the role of conceptual analysis in the acquisition of knowledge, the link between conceivability and necessity and the nature of Fregean sense, secondary qualities, narrow content, meaning, and communication. Whether these claims of illumination are warranted depends on which, if any, of the various interpretations of the 2D framework are coherent. This is a contested matter. In this survey we will describe the logical machinery and outline its deployment by a few philosophers, emphasizing possible connections to the contingent *apriori* and necessary empirical.

Logical Machinery

However contested its various interpretations, the technical machinery required for a 2D logic of the kind described above is relatively straightforward. We may take a 2D *model* to be a pair M = (W,V), where W is a nonempty set (the possible worlds) and V is a function (the valuation) that assigns a subset of W × W to each sentence letter. A sentence letter p is true in M at a pair (u,v) of possible worlds (u-world and e-world, respectively) if (u,v)ΣV(p). We may consider languages with a variety of connectives and constants, including those with the following clauses defining A is true at (u,v) in M (written $(u,v) \models_M A$). (The quantifiers here are all taken to range over W.)

$(u,v) \models_M \Box A$	iff	$\forall u' \forall v' \; (u',v') \models_M A$
$(u,v) \models_M \boxed{1} A$	iff	$\forall v' \; (u,v') \models_M A$
$(u,v) \models_M \boxed{2} A$	iff	$\forall u' \; (u',v) \models_M A$
$(u,v) \models_M ① A$	iff	$(u,u) \models_M A$
$(u,v) \models_M ② A$	iff	$(v,v) \models_M A$
$(u,v) \models_M \otimes A$	iff	$(v,u) \models_M A$
$(u,v) \models_M \boxed{\Box} A$	iff	$\forall w \; (w,w) \models_M A$
$(u,v) \models_M \boxed{\cdot} A$	iff	$\forall w \; (v,w) \models_M A$
$(u,v) \models_M \boxdot A$	iff	$\forall w \; (w,u) \models_M A$
$(u,v) \models_M \mathbf{I}$	iff	$u = v$

We may say that a formula A is *valid* if, for any model M and any world u of M $(u,u) \models_M A$, and that A is *strongly valid* if, for any M and any worlds u,v of M $(u,v) \models_{MA}$. (Motivation for the definitions of validity will be suggested in subsequent sections.)

These connectives (or their informal counterparts) have appeared in the logic and philosophy literatures on two-dimensionalism under a variety of names and notations. The nomenclature above follows a convention once suggested by Brian Chellas. Connectives

roughly analogous to the standard necessity operator are represented by something whose outer boundary is a square. Duals of these connectives are named by rotating the boundary 45 degrees to to obtain a diamond. For example, ◇, the dual of □, is a connective having the truth conditions: $(u,v) \models_M \Diamond A$ iff $\exists w\ (w,w) \models_M A$. If a connective **C** is self dual (so that **C**A always takes the same truth value as $\neg \mathbf{C} \neg A$) then it is represented by something whose boundary (in the cases above, a circle) remains the same when so rotated. Beyond these conventions, numbers appearing within the connective indicate the coordinate on which the truth of a formula made by applying the connective depends. ⚀ and ⚁ are intended to evoke the image of dominos. If their truth clauses are written in terms of accessibility relations those relations, (t,u)R(v,w) iff u = v, and (t,u)R(v,w), iff t = w are the relations that permit dominos to be placed to the right and left of others in play. It is common to view the pairs of worlds as forming an array in which (for reasons unclear) the vertical distance from the top determines the u-world and the horizontal distance from the left indicates the v-world. On this understanding, the identity pairs (u,u) form a *diagonal* from upper left to lower right. ①A, ②A and A are true at a pair if A is true throughout the corresponding row, the corresponding column, and the diagonal, and the other connectives can likewise be understood in geometric terms.

A logical system based on the connectives □, ①, ②, ⓵, ⓶, and ⊗ (and the notion of strong validity), referred to as the *basic* two-dimensional system ***B***, is axiomatized in Segerberg 1973. Systems based on ⚀ and ⚁ are investigated in Kuhn 1989 and Venema 1992.

These connectives are interdefinable in various ways. For example, each square and diamond connective is interdefinable with its dual.

□A may defined as ① ②A, ②① A, ⚁⚁ A, or ⚀⚀ A;
□A, as ② ⓵ A or ② ⓵A;
②A, as ⊗⓵A or ◇(**I**∧A);
⓵A, as ⊗②A or ◇ (**I**∧A);
②A, as ⊗①⊗A or ⓶⚁ **A**;
① A, as ⊗②⊗A or ⓵⚀A;
⚀A, as ⓶①A or ⊗①A;
⚁ A, as ⓵②A or ⊗②A.

Thus all of the above connectives except **I** are definable in ***B*** and in fact they can all be defined in a language that contains just ①, ⓵, and ⊗ as primitive. Adding ② and ⓶ to these three connectives allows us to obtain a useful normal form result. To see this, note that the following biconditionals are valid (where each i and j inside a circle or square is either 1 or 2, ĩ is the "opposite" number as i and ○ may be ⓵, ⓶ or ⊗:

ⓘⓙA ↔ ⓘA
ⓘ⊗A ↔ ⓘA
ⓘ①A ↔ ①A
ⓘ①̃A ↔ ①⊗A
⊗ⓘ A ↔ ĩA
⊗⊗A ↔ A
⊗①A ↔ ①̃⊗A
○(A∨B) ↔ ○ A∨○B
○¬A ↔ ¬○A

By means of these biconditionals, we can push every circular connective occurrence inward, past occurrences of squares and boolean connectives, shortening all strings of circular connectives to length at most one, until each occurrence of ①, ②and ⊗ governs only a sentence letter. The observation that all formulas can be written in this "inner circle" form facilitates a proof that the basic two-dimensional logic ***B*** is equivalent to the fragment, PL(x,y), of ordinary predicate logic that contains only binary predicate letters and the two variables x and y. Note that the 2D models M = (W,V) can be viewed as models for dyadic predicate logic. W becomes the *domain* of the model and the set of pairs that V assigns to the i'th sentence letter $\mathbf{p}_i$ becomes the extension of the i'th predicate, $\mathbf{P}_i$**.** A translation t that maps the inner circle formulas one–one onto PL(x,y) can be defined as follows.

$t(\mathbf{p}_i) = \mathbf{P}_i xy$
$t(①\mathbf{p}_i) = \mathbf{P}_i xx$
$t(②\mathbf{p}_i) = \mathbf{P}_i yy$
$t(⊗\mathbf{p}_i) = \mathbf{P}_i yx$
$t(A \vee B) = t(A) \vee t(B)$
$t(\neg A) = \neg t(A)$
$t(\boxed{1}A) = \forall y\, t(A)$
$t(\boxed{2}A) = \forall x\, t(A)$

We can now establish by formula induction that truth of an inner-circle formula A at the pair (u,v) corresponds to truth of the predicate logic formula t(A) when u and v are assigned to x and y. In symbols, $(u,v) \models_M A$ iff $M \models t(A)[u,v]$ (and consequently, for all φ in PL(x,y), $M \models [u,v]$ iff $(u,v) \models_M t^{-1}(\varphi)$.) Thus, the inner circle formulas and the formulas of PL(x,y) are essentially the same and, since any formula with connectives above other than **I** is equivalent to an inner-circle formula, the full two-dimensional logic without **I** is essentially PL(x,y). It is perhaps not as obvious as it may seem that every two-dimensional formula can be expressed with two variables. The formula $◈◈(p_1 \wedge ◈(p_2 \wedge ◈p_3))$, for example, might naturally be associated with the four-variable formula $\exists x \exists y \exists z \exists w(P_1xy \wedge P_2yz \wedge P_3zw)$. By the construction above, however, it can be seen to be equivalent to $\exists x(\exists y(P_1xy \wedge \exists x(P_2yx \wedge \exists\, y(P_3xy)))$. Our translation can be extended to include **I** by setting t(**I**) to (x = y) (and t(①**I**), t(②**I**), t(⊗**I**) to (x = x),(y = y) and (y = x)), thereby establishing the equivalence of the full system to the version of PL(x,y) with identity, which we might write $PL^{=}(x,y)$. Since identity is not definable in PL(x,y), it follows that **I** is not definable in ***B***. None of the various flavors of two-dimensionalism make use of all the connectives described above, but most can be understood as built upon some of them, and hence as built upon some fragment of $PL^{=}(x,y)$.

Davies and Humberstone

The application in which the technical apparatus sketched above lies closest to the surface is contained in the influential Davies and Humberstone 1980,where it is used to explicate a distinction drawn in Evans 1979 between "deep" and "superficial" necessity. $(u,v) \models_M A$ is read "A is true *at* v from the perspective of u as the actual world." 1 and ① (written there as □ and **A**) are read *necessarily* and *actually*. In a language with just these two connectives, the truth value of A at (u,v) depends only on the truth values of its subformulas at pairs (u,v′). u "stores" the world regarded as actual so that the *actually*

operator can refer to it even when that operator lies within the scope of necessity operators. While this may be a plausible characterization of the semantic roles of *actually* and *necessarily*, it has the peculiar consequence that $@A \leftrightarrow \Box@A$ is strongly valid, i.e., that *actually* A is equivalent to *necessarily actually* A.

If, at v, A is true at u from the perspective of u as actual then, at any w, A is true at u from the perspective of u as actual. Davies and Humberstone diagnose the apparent peculiarity as stemming from the thought that "another world might have been actual," and suggest $\Box_2$ (with the reading *fixedly*) as a way of expressing that a formula is true no matter what world is taken as actual. The reader puzzled by the previous formula is confusing it with $@A \leftrightarrow \Box_2@A$, which is not valid (even in the weaker sense). Furthermore, we can understand the right side of this biconditional, or its equivalent $\blacksquare A$ (read here as *fixedly actually* A), to represent something close to Evans' *deep necessity*, and $\Box_1 A$ to represent his *superficial necessity*. Evans's examples of these notions employed the device (which may or may not be exemplified in natural language) of *descriptive names*. A descriptive name for x is a name that behaves semantically like the definite description *the object that is actually P*, where P is a predicate that holds only of x and is *actually* interpreted as $@$. Thus descriptive names are rigid (i.e., they have fixed denotations) with respect to e-worlds, but not u-worlds, whereas ordinary proper names are said to be rigid with respect to both worlds and ordinary definite descriptions (for example, "the tallest woman in the room,") may be rigid with respect to neither. We may call a description formed with the *actually* operator in this way a *rigidified* description. If λ is the descriptive name associated with *the actual length of the standard meter bar*, then the sentence *if there is a unique length of the standard meter bar, then λ is the length of the standard meter bar* is deeply but not superficially necessary. The sentence *if there is a unique length of the standard meter bar, then the length of the standard meter bar is the length of the standard meter bar*, however, is both superficially and deeply necessary. Within Davies' and Humberstone's framework, we can understand Evans's distinction as saying that the first sentence is true along the diagonal, i.e., "whichever world had been actual, [the sentence] would have been true at that world considered as actual," while the second is true along the horizontal, i.e., from the perspective of the actual world as actual, the sentence is true at every world. In logical notation, A is deeply necessary with respect to pair (u,v) and model M if $(u,v) \models_M \blacksquare A$ and superficially necessary with respect to (u,v) and M if $(u,v) \models_M \Box_1 A$. Since $\blacksquare A$ is true at one pair iff it is true at all, can we take deep necessity to be a property of sentences relative to models. We can similarly take superficial necessity to be a property of sentences relative to models and u-worlds. In applying these notions to natural language, we understand superficial necessity *simpliciter* as superficial necessity with respect to the actual world.

The informed reader will have already noted the similarity between the example here of a deeply, but not superficially, necessary sentence and the example in Kripke 1972 of a sentence expressing a contingent truth knowable *apriori*. Indeed, once we see how the example is constructed, the framework of Davies and Humberstone allows the construction of even simpler "toy"examples of the phenomenon. $\blacksquare(P \leftrightarrow @P)$ is strongly valid, but $\Box_1 (P \leftrightarrow @P)$ is not even weakly valid, so sentences of the form $(P \leftrightarrow @P)$ are all deeply, but not superficially, necessary sentences. Indeed, Davies and Humberstone tell us, "one can know *apriori* that grass is actually green iff grass is green." Kripke's and Evans's examples can be replaced by ones requiring no special views about the semantics of proper names. Davies and Humberstone are cautious about the degree to which deep necessity and *apriority* coincide. If true proper names are rigid with respect to u-worlds and e-worlds, then true identity and difference statements employing them must be

deeply necessary, yet a plausible case can be made that some of these are not knowable without observation. In the other direction, other interpretations of the 2D framework allow arguments that sentences true along the diagonal are knowable *apriori*, but Davies and Humberstone are content to remark that they have yet to find an example representable in their formal language that is not. Subsequent work has shown less restraint.

In a similar way, we can construct toy examples of the necessary empirical. Since *Grass is green* is true at (u_0,u_0) where u_0 represents the actual world, *grass is actually green* is true at (u_0,v) for every possible world v, and so it is superficially, but not deeply, necessary. Since there is no way to know that this sentence is true without knowing that grass is green, we have a simple example of a necessary truth, not knowable *apriori*. Davies and Humberstone consider, without endorsement, suggestions that philosophically more interesting examples might have similar structure. *Water is* H_2O (for chemically naive speakers). *Red objects reflect light of wavelength about 650nm. Actions possessing a property that actually arouses in me a feeling of disapproval are wrong.* The subject of the sentence in each case can be regarded as containing an explicit or implicit *actuality* operator. So understood, each may be seen as superficially, but not deeply necessary, and as necessary but not knowable *apriori*.

When a sentence is asserted, we may take it as asserting the truth of the sentence at the actual world from the perspective of that world considered as actual. If the connectives of 2D logic represent logical form, the two notions of validity correspond to two common characterizations of logical truth. A sentence is valid if it is true (i.e., it can be truly asserted) in virtue of its logical form. It is strongly valid if it is necessarily true in virtue of its logical form. Thus, this interpretation of the 2D framework, if correct, would show that the two characterizations come apart. Another useful distinction, somewhat related, is made by Davies and Humberstone (with attribution to Michael Dummett) and emphasized in Lewis 1981. Considered as assertions, our two sentences about the meter bar and, more generally, two sentences of the form P and $\textcircled{1}$P, apparently say the same thing. Yet each contributes differently the truth of sentences containing them. For $\boxed{1}$ (grass is green) is false, while $\boxed{1}\textcircled{1}$ (grass is green) is true. We may say that A and B have the same *assertive content* (relative to M) if they agree on the diagonal, i.e., if $\models_M \Box(A \leftrightarrow B)$. They have the same *ingredient sense*, if they agree at all pairs, i.e., If $\models_M \boxed{1}\boxed{2}(A \leftrightarrow B)$.

Although Davies and Humberstone are relatively cautious in their claims on behalf of two-dimensionalism, it may be useful to list a few questions about their interpretation of the logical machinery. Analogous questions often apply to interpretations discussed in subsequent sections.

1. Is the interpretation of $(u,v)\models_M A$ (and the role of the u-world in particular) coherent? How should we understand locutions like *true at v from the perspective of u considered as actual*, and *if the actual world were u*. It is not obvious that these come to the same thing or that the second describes an entertainable condition. Perhaps the motivating "intuition" that another world might have been actual is confused and ought to be discouraged rather than represented.
2. Is the set of u-worlds and the set of e-worlds really the same? In answering question one, we may determine that not every possible world might have been actual or that the objects we might "consider" as possible worlds are not possible worlds at all. (See discussion of Chalmers, below.) We then lose the homogeneity characteristic of 2D logic that made talk of the diagonal sensible.

3. Are the explanations of the necessary-empirical and the contingent-*apriori* plausible? In the Davies–Humberstone framework, the idea underpinning these explanations is that these sentences contain, explicitly or implicitly, something like an *actuality* operator. Notice also that the puzzling pairs of properties are attributed to sentences,whereas Kripke spoke of *statements*.

Kaplan

David Kaplan is routinely cited as a father of two-dimensionalism, but the logical machinery employed in Kaplan 1978 and 1989, and its motivation and interpretation, are quite distinct from those described in other sections of this survey (and he seems anxious to distance himself from the enterprise –1989, p 512). The main concern in the Kaplan papers is the logical and semantic properties of *indexical* expressions like *I, here*, *now* and the lessons that these might have for nature of meaning generally. In particular, Kaplan observes that the outlook outlined in the first paragraph here, whereby traditional modal logic is generalized by simply replacing possible worlds with n-tuples of parameters, will fail to capture the sense in which a sentence like *I am here now*, is deeply and universally true. If we restrict the n-tuples to the "proper" ones, say the 4-tuples (w, a, p, t) where agent a in world w is producing a linguistic expression at place p and time t, we will falsely represent the sentence as necessarily true. If, on the other hand, we allow the parameters to vary independently, we will improperly suggest that it can be uttered falsely. The solution is to recognize that the parameters here can play two roles. They can be features of *context* in which an expression is produced or the *circumstances* in which the content expressed by an expression in such a context is evaluated. The content of an utterance of *I am here* by David Kaplan in Portland (in the actual world) on March 26, 1977, for example, is the proposition that Kaplan was in Portland then. Since this proposition is false when evaluated in circumstances where he was elsewhere on that date, it is not necessary. Contents are supposed to be "what is said" by an expression in a context. They can be conveniently represented by functions from circumstances to extensions, though we should remember that such functions are not identical to contents they represent. We can similarly represent another variety of meaning, which Kaplan calls *character* by functions from contexts to contents.

If we take possible worlds as the sole features of context and circumstance, the logic of demonstratives in Kaplan 1978 is built on the 2D logic described here with connectives ① (*actually*) and 🄸 (*necessarily*) and validity understood in the weaker sense. Because Kaplan's version is formulated within predicate, rather than propositional, logic, however, he is able to introduce a formal device by which the rigidified descriptions of Davies and Humberstone (and therefore the descriptive names of Evans) can be represented. For any term t, *dthat* t, is a term whose extension at (u,v) is the extension of t at (u,u). Thus, if **s** represents *the length of the standard meter bar*, then *dthat* s represents the actual length of the standard meter bar. What Kaplan's logic obviously *omits* is anything like the connectives 🄿 (*fixedly*) and 🅐 (*fixedly actually*), that feature so prominently in the discussion of Davis and Humberstone. If the interpretation of ○**A** at a pair (u,v) depends on the interpretation of A at pairs (u′,w) for u ≠ u′, then ○ is operating on the character of A rather than merely its content. Kaplan 1989 labels such operators, and the expressions that they might represent, *monsters*, and maintains that English (unless being used metalinguistically) has no such devices. Although Davies and Humberstone's *fixedly* is not an expression of ordinary English and Kaplan's *I*, *here* and *now*

are not monsters, the basis for the general claim is somewhat unclear and it has been questioned by a number of authors. (See, for example, Schlenker 2003.)

Kaplan does not emphasize the connections between this framework and the contingent *apriori* or the necessary empirical. Kaplan 1978 (p. 85) alludes to a "structural" similarity between his character/content distinction and Kripke's *apriori*/necessary one. Kaplan 1989 (p.550) mentions examples reminiscent of those in Davies and Humberstone. If s and t are definite**,** nonrigid descriptions for which s = t is a contingent, empirical truth, then *dthat* s = *dthat* t is both necessary (because all identity statements with rigid designators are necessary) and empirical (because it is true iff s = t). On the other hand, s = *dthat* s is both contingent (because s is nonrigid) and knowable *apriori* (because it is valid in the logic of demonstratives). The argument for this last claim would seem to indicate that the contingent *apriori* extends beyond the cases mentioned by Kripke to include examples like Kaplan's *I am here now*. Kaplan nowhere makes such an observation, perhaps because the considerations raised in question three of the previous section are so salient for these examples. *I am here now* is valid in Kaplan's logic, and so we can know *apriori* that it is true. Its content in a given u-world, however, is something empirical, like the proposition that Kaplan was in Portland in 1977. While we might know *apriori* that the sentence *I am here now* is true, it requires empirical knowledge to know what proposition the sentence expresses in a particular context. Soames 2005 maintains that all of Kaplan's examples of puzzling pairs (but not Kripke's) are defective because the proposition alleged to be *apriori* is not the same as that alleged to be empirical (and that the citations above represent a misleading and unfortunate digression in Kaplan's otherwise insightful observations). Lewis 1994, and other of what Soames calls "strong" two-dimensionalists, however, maintain that all alleged examples of puzzling pairs have this feature.

Lewis

In an earlier paper (1981), Lewis suggested that the idea that there is some intuitive, univocal notion of "what is said" captured by Kaplan's content is illusory. On the Lewis 1981 version of two-dimensionalism, the appropriate distinction is not between features of context and features of circumstance but rather between context as a whole and a small package (or *index*) of contextual features that are independently shiftable. Every sentence gets a truth value at a context, but, for a compound sentence, that truth value depends on the truth values of its subsentences at indices in which certain contextual features have been shifted (and of the mode of compounding in question). *I have been here* is a compound of the sentence *I am here* using the *it has been that* mode. The truth value of an utterance of *I have been here* in a context with Kaplan as speaker in Portland in 1977 is determined by the truth values of *I am here* at indices in which the time coordinate of that context is shifted toward the past. After the coordinate has shifted, the index may correspond to no context at all. (Kaplan may not be there then.) Since any sentence may occur either alone or as part of a compound, we must evaluate all sentences relative to both contexts and indices. These are the two dimensions. From this perspective, the absence of monsters is a matter of stipulation. If there is any operator ○ for which the truth of ○A in a context c depends on the truth of A when some feature of c has shifted then that feature is, by definition, part of the index. This does not rule out the possibility that two different operators might make use of the same contextual feature. In that case indices may require two coordinates to store different

values of that feature. Perhaps this situation would be as monstrous as the presence of the context-altering creatures that frightened Kaplan.

Stalnaker

For Stalnaker, the need for 2D logic arises in connection with an explanation of *assertion*. We start with the familiar idea that the content of an assertion, a proposition, can be roughly identified with a set of possible worlds (the worlds in which the proposition is true). Each participant brings to a conversation a set of *presuppositions*, i.e., a set of propositions p such that he is disposed to act as if he takes p to be true, takes the others to take it to be true as well, takes them to take him to take it to be true, and so on. The intersection of these propositions, the speaker's *context set*, will comprise all worlds that the speaker takes to be "live options" in the conversation. The participants in a conversation value communication. Since communication is impeded when presuppositions diverge, context sets will tend to become *non-defective*, i.e., identical for all. Given a nondefective context, conversants have a further common interest in narrowing it, that is, in more fully specifying the way the world is taken to be, though if they disagree about the facts their interests in how this should be done may conflict. An assertion of the proposition p is, in essence, a proposal to narrow the context set by eliminating all worlds where p is false (or, equivalently, by selecting just those making p true).

One difficulty with this very plausible picture is that necessary propositions are nowhere false, and so cannot narrow any context set. Yet utterances like *Hesperus is identical to Phosphorus, it is now three o'clock*, and an *ophthamologist is an eye doctor* appear to be respectable assertions of just such propositions. These appearances, Stalnaker contends, are deceiving. *Hesperus is Phosphorus* does *standardly* express a necessary proposition. Since asserting a necessary proposition fails to narrow the context set, however, it would violate a fundamental principle of conversation. We are therefore justified in interpreting one who appears to be doing so as asserting another instead. (Stalnaker is here invoking ideas of Paul Grice. See Chapter 7.7 of this volume.) That *Hesperus is Phosphorus* expresses a necessary proposition follows from the fact that *Hesperus* and *Phosphorus* rigidly denote the planets visible in the evening and morning sky and that these are one and the same. The speaker who appears to assert that Hesperus is Phosphorus, it is reasonable to suppose, fails to presuppose this last condition. His context set includes a world w in which the morning and evening appearances are of a single planet and a world w′ in which they are of different planets. In w his utterance expresses a necessary truth; in w′, a necessary falsehood. The situation can be modeled in the now familiar way by a 2D model M, where $(u,v) \models_M A$ indicates that what A expresses in u is true in v. If the truth values of A are arranged geometrically as described above, the row of truth values at vertical position u represents the proposition expressed by A in u. The speaker is proposing to eliminate w′, but not w, from the context set. He cannot do this by asserting propositions that would be expressed by A or ¬A in w or w′, for each of those would eliminate either nothing or everything. He *can* do it by asserting the proposition expressed in either world by ①A (†A in Stalnaker's notation). ①A can be regarded as expressing, at every world, that the proposition expressed by *Hesperus is Phosphorus* is true, or that the denotations of *Hesperus* and *Phosphorus* are identical. In general, it is often reasonable to interpret what appears to be an assertion of a necessarily true proposition that A as instead an assertion of a contingent proposition that ①A.

Stalnaker 2004 emphasizes that he is not suggesting a semantical theory under which the proper name *Hesperus*, like Kaplan's indexicals *I* and *here*, gets different interpretations in different contexts. Semantical considerations still determine that *Hesperus is Phosphorus* standardly expresses either a necessarily true proposition or a necessarily false one, and empirical considerations help us determine that it is the former. Pragmatic considerations that presuppose and utilize the standard interpretation, however, may lead us to conclude that an utterance of that sentence is an assertion of an entirely different, contingent proposition. Utterances of *Hesperus* do denote different individuals in different u-worlds, but this should not be thought of as reflecting the meaning of *Hesperus*. Under appropriate conditions the context set might contain a world where such an utterance denotes a horse or a real number.

Similar considerations suggest that Stalnaker's two-dimensionalism, though perhaps explaining how sentences that standardly expresses necessary propositions can convey information about the world, has little to tell us about the existence and nature of the contingent *apriori* or the necessary empirical. Stalnaker 1981 remarks that $\Box$ can be understood as an *apriori* truth operator, but Stalnaker 2004 rightly expresses strong misgivings. If it is sometimes appropriate to interpret *Hesperus is Phosphorus* as expressing something necessarily false, it may be similarly appropriate to do so for *7+5 = 12*, *Tuesday follows Monday*, or any other standardly *apriori* sentence. If so, then these will be false somewhere on the diagonal and an application of $\Box$ will produce sentences everywhere false.

Chalmers

The currency of the term *two-dimensionalism* is largely due to David Chalmers, the subject's most voluble and energetic expositor. Yet, on Chalmers' preferred interpretation, 2D logic, as described here, is largely hidden from view. Having climbed the 2D ladder, Chalmers is happy to put it aside. (Some qualification is warranted regarding this and other remarks in this section. A large proportion of Chalmers' writings on two-dimensionalsm is devoted to interpretations other than what I call his preferred interpretation. "Semantic pluralism" allows him to regard some of these as viable alternatives, capturing notions of meaning distinct from his own primary concern. Others are presented as reformulations of the preferred framework, taking different notions as primitive than the preferred framework but, under appropriate philosophical assumptions, arguably equivalent to it. The intention is to avoid losing the reader who might have scruples about his choice of primitives. My remarks concern the preferred interpretation. A further warning: in this section more than others, qualifications and refinements are omitted to convey basic ideas in reasonable space.)

Chalmers sees two-dimensionalism as clarifying connections between cognitive content and necessity revealed to us by Frege and Carnap's discussions of *sense* and *intension*, but subsequently obscured by Kripke's persuasive examples of necessary empirical statements. For this reason, he deliberately seeks an interpretation satisfying a condition like the one Stalnaker repudiated. He takes u-worlds to be *scenarios*, rather than possible worlds, corresponding to "ways that the world might turn out to be, for all we know *apriori*." On the preferred analysis, scenarios are identified with equivalence classes of complete sentences in some idealized (infinitary) language capable of representing every hypothesis about the world. A complete sentence is one that is epistemically possible (this notion is primitive) and implies S or $\neg$S for every sentence S in the

language. Sentences within equivalence classes all imply each other. Some scenarios imply ("verify") utterances of *Hesperus is Phosphorus*, others verify utterances of *Hesperus is not Phosphorus*, but none verifies the conjunction of these and none verifies *Venus is not Venus*. A scenario verifying *Hesperus is not Phosphorus* might have that sentence as a conjunct, but it might instead have *the bright celestial objects visible in the morning and evening are Mars and Mercury, respectively*, or perhaps *no celestial object appears in both the morning and evening*. In general, Chalmers argues, for any scenario u, there will be a "relatively limited" vocabulary V such that if u verifies an utterance S, then u verifies some sentence D containing only expressions of V that implies S. There is no requirement that expressions of V be qualitative and no requirement that objects mentioned in s be *definable* in V.

Since scenarios and worlds are distinct, this conception affords no notion of a *diagonal*. Instead, Chalmers defines the "1-intension" of a sentence S to be the function that assigns to each scenario u the value T if u implies S and F if it implies ¬S. It now becomes plausible to argue that S is *apriori* if and only if its 1-intension always takes the value T. Though it is unclear that his preferred interpretation requires it, Chalmers argues that there is a *correspondence* between scenarios and worlds (or *centered* worlds, i.e., worlds with time, place, and speaker marked) so that the 1-intension corresponds to a diagonal after all, but he insists that it should not be seen as deriving from it. The "2-intension" is just the familiar function from possible worlds to truth values. Thus, sentences are not (or not fundamentally) evaluated at scenario-world pairs, but at worlds and scenarios alone. The intuition is that an utterance of S is *verified* by *scenario* u if it is true if the actual world turns out to really *be* as described by u. It is *satisfied* by *possible world* v if it would be true if the actual world *were* as described by v. The scenario according to which no body is visible morning and evening verifies *Hesperus is not Phosphorus*, but the world with no such body still satisfies *Hesperus is Phosphorus*. If the interpretation has the properties it was designed to, then all the puzzling pair examples have similar structures. An utterance is true *apriori* if and only if its 1-intension takes the value T at all scenarios and it is necessary if and only if its 2-intension takes the value T at all worlds. It should not be surprising that each of these conditions can obtain without the other. Whether that implies that there are necessary empirical and contingent *apriori* propositions depends on whether one takes the 1-intension and 2-intension of an utterance to correspond to two propositions, both expressed by that utterance, or whether one takes them to be parts of a single proposition expressed by it. Chalmers leans toward the latter. To say that puzzling-pair examples have epistemological and metaphysical properties in virtue of different aspects of their meanings does seem more reasonable than to insist they express two propositions.

References

Chalmers, D. (2006) "The foundations of two-dimensional semantics," in M. Garcia-Carpintero & J. Macia, eds., *Two-Dimensional Semantics: Foundations and Applications*, New York: Oxford University Press, 2006.

—— (2006a) "Two-dimensional semantics," in *Oxford Companion to the Philosophy of Language*, E. Lepore and B. Smith, eds., New York: Oxford University Press.

—— (forthcoming) "The nature of epistemic space," in A. Egan and B. Weatherson, eds., *Epistemic Modality*. Oxford: Oxford University Press.

Davies, M. and L. Humberstone (1980) "Two notions of necessity," *Philosophical Studies*, 38:1–30.

Evans, M. (1979) "Reference and contingency," *The Monist* 62: 161–189.

Humberstone, L. (2004) "Two-dimensional adventures," *Philosophical Studies* 118: 17–65.

Jackson, F. (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Kaplan, D. (1978) "On the logic of demonstratives," *Journal of Philosophical Logic*, 8:81–98.

—— (1989) "Demonstratives," in J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*, 481–563, Oxford: Oxford University Press.

Kripke, S. (1972) "Naming and necessity," in D. Davidson and G. Harman, eds., *Semantics of Natural Language*, 253–355, Dordrecht: D. Reidel and Company.

Kuhn, S. (1989) "The domino relation: Flattening a two-dimensional logic," *Journal of Philosophical Logic*, 18(2):173–195.

Lewis, D. (1981) "Index, context, and content," in S. Kanger and S. Ohman, eds., *Philosophy and Grammar: Papers on the Occasion of the Quincentennial of Uppsala University*, 79–100, Dordrecht: D. Reidel Publishing Co.

—— (1994) "Reduction of mind," in S.Guttenplan, ed., *Companion to the Philosophy of Mind*, Oxford: Blackwell.

Reichenbach, H. (1947) *Elements of Symbolic Logic*, New York: Random House.

Schlenker, P. (2003) "A Plea for Monsters,' *Linguistics and Philosophy*, 26(1): 29–120.

Segerberg, K. (1973) "Two-dimensional modal logic," *Journal of Philosophical Logic*, 2(1):77–96.

Stalnaker, R. (1981) "Assertion," in P. Cole, ed., *Radical Pragmatics*, New York: Academic Press.

—— (2004) "Assertion Revised: On the interpretation of two-dimensional modal semantics," *Philosophical Studies*, 118(1–2) 299–322.

Thomason, R. (1984) "Combinations of tense and modality," in D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, 135–165. Dordrecht: D. Reidel Publishing Co.

Venema, Y. (1992), "A Note on the Tense Logic of Dominoes,' *Journal of Philosophical Logic* 21(2):173–182.

Further Reading

Garcia-Carpintero, M. and J. Macia, eds. (2006) *Two-Dimensional Semantics*, Oxford: Oxford University Press. (An anthology including interpretations, applications, and criticism of the two-dimensional framework. The contributions of Davies, Evans, and Stalnaker to Stoljar and Davies (2004), are reprinted, and it also contains Chalmers 2006 and eleven other papers.)

Marx, M. and Y. Venema (1997) *Multi-Dimensional Modal Logic*, Kluwer, Dordrecht Academic Publishers. (The only book-length treatment of many-dimensional logics. Contains a chapter on two-dimensional logic with a wealth of technical results on logical systems with connectives like those described here—but not the normal form and translation results sketched above.)

Soames, S. (2005) *Reference and Description: The Case Against Two-Dimensionalism*, Princeton: Princeton University Press. (A detailed and sustained attack on two-dimensionalism, which the author sees as a misguided attempt to revive Fregean descriptivism in the face of compelling arguments by Saul Kripke against it. If Chalmers is the subject's most energetic expositor, Soames is its most energetic critic.)

Spencer, C. (2010) "Two-dimensional semantics," *Oxford Bibliographies Online—Philosophy*. (A richly annotated, extensive and current bibliography.)

Stoljar, D. and M. Davies, eds. (2004), *Philosophical Studies* 118(1–2). A double issue of the journal containing material from a 2002 conference. In addition to Humberstone 2004 and Stalnaker 2004, it contains a useful introduction by the editors, a shortened version of Chalmers 2006, papers by David Braddon-Mitchell, Martin Davies, Frank Jackson, Frederick Kroon, Michaelis Michael, Philip Pettit, Laura Schroeter, and a posthumous publication of Gareth Evans's comments to Davies and Humberstone concerning their 1980 paper.

5.6 MANY-VALUED LOGICS

Nicholas J.J. Smith

1 Introduction

A many-valued (aka multiple- or multi-valued) *semantics*, in the strict sense, is one which employs more than two truth values; in the loose sense it is one which countenances more than two truth statuses. So if, for example, we say that there are only two truth values—True and False—but allow that as well as *possessing the value True* and *possessing the value False*, propositions may also have a third truth status—*possessing neither truth value*—then we have a many-valued semantics in the loose but not the strict sense. A many-valued *logic* is one which arises from a many-valued semantics and does not also arise from any two-valued semantics [Malinowski, 1993, 30]. By a 'logic' here we mean either a set of *tautologies*, or a *consequence relation*. We can best explain these ideas by considering the case of classical propositional logic. The language contains the usual basic symbols (propositional constants p, q, r, . . .; connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$; and parentheses) and well-formed formulas are defined in the standard way. With the language thus specified—as a set of well-formed formulas—its semantics is then given in three parts. (i) A *model* of a logical language consists in a free assignment of semantic values to basic items of the non-logical vocabulary. Here the basic items of the non-logical vocabulary are the propositional constants. The appropriate kind of semantic value for a proposition is a *truth* value, and so a model of the language consists in a free assignment of truth values to basic propositions. Two truth values are countenanced: 1 (representing truth) and 0 (representing falsity). (ii) Rules are presented which determine a truth value for *every* proposition of the language, given a model. The most common way of presenting these rules is via truth tables (see Figure 5.6.1). Another way of stating such rules—which will be useful below—is first to introduce functions on the truth val-

α	β	$\neg\alpha$	$\alpha \wedge \beta$	$\alpha \vee \beta$	$\alpha \rightarrow \beta$	$\alpha \leftrightarrow \beta$
1	1	0	1	1	1	1
1	0		0	1	0	0
0	1	1	0	1	1	0
0	0		0	0	1	1

Figure 5.6.1 Classical truth tables

ues themselves: a unary function $\overset{\star}{\neg}$ and four binary functions $\overset{\star}{\wedge}$, $\overset{\star}{\vee}$, $\overset{\star}{\rightarrow}$ and $\overset{\star}{\leftrightarrow}$ (see Figure 5.6.2). Representing the truth value of α (on a given model) as $[\alpha]$, we then specify the truth values of compound formulas as in Figure 5.6.3. Once one becomes familiar with the distinction between connectives and truth functions, it is customary to use the same symbols for both and to let context disambiguate. As it generally increases readability, I shall mostly follow this practice below (i.e. omit the $\star$'s on truth functions). (iii) Definitions of *tautology* and *logical consequence* are introduced. In this case, a tautology is a proposition which gets the value 1 on every model (e.g. $p \vee \neg p$, $p \rightarrow p$), and a proposition α is a logical consequence of the set of propositions Γ (written $\Gamma \models \alpha$) if, on every model on which every proposition in Γ has the value 1, α has the value 1 (e.g. $\{p, p \rightarrow q\} \models q$, $\{p\} \models p \vee q$). Classical *logic* is then the language just introduced together with either the set of tautologies, or the consequence relation, just defined. The definition of a logic in terms of a consequence relation is more powerful, in that once we have the consequence relation, we can reconstruct the set of tautologies as the set of propositions α such that $\emptyset \models \alpha$. However sometimes we are interested only in tautologies—hence we allow that a logic may be specified just by giving a set of tautologies, without a consequence relation.

As we have just seen, a logic can be specified as the one which arises from a certain semantics. Logics can also be specified in terms of proofs. Proofs come in many different forms. Consider, for example, axiomatic proof systems. Finitely many propositions are taken as axioms. One or more rules of inference are specified: they take one or more propositions as input, and give a proposition as output. A formal proof (in a given axiomatic system) is then defined as a finite list of propositions, each of which is either an axiom, or follows from earlier propositions in the list by a rule of inference. A proposition α is said to be a theorem of the system if there is a proof whose last line is α. There are many known axiom systems which have the property that their theorems are exactly the tautologies of classical logic. Thus, classical logic is said to be finitely axiomatizable. We can also define consequence relations using axiomatic proof procedures, by allowing assumptions in proofs. So axiomatic proof systems give us an alternative way of characterizing logics.

		$\overset{\star}{\neg}$	$\overset{\star}{\wedge}$	$\overset{\star}{\vee}$	$\overset{\star}{\rightarrow}$	$\overset{\star}{\leftrightarrow}$
1	1	0	1	1	1	1
1	0		0	1	0	0
0	1	1	0	1	1	0
0	0		0	0	1	1

Figure 5.6.2 Classical truth functions

$$
\begin{aligned}
[\neg\alpha] &= \overset{\star}{\neg}\,[\alpha] \\
[\alpha \wedge \beta] &= [\alpha] \overset{\star}{\wedge} [\beta] \\
[\alpha \vee \beta] &= [\alpha] \overset{\star}{\vee} [\beta] \\
[\alpha \rightarrow \beta] &= [\alpha] \overset{\star}{\rightarrow} [\beta] \\
[\alpha \leftrightarrow \beta] &= [\alpha] \overset{\star}{\leftrightarrow} [\beta]
\end{aligned}
$$

Figure 5.6.3 Rules for assigning truth values to compound propositions

In §2 we look at systems of many-valued semantics and their associated logics and in §3 we mention some of the uses to which these systems have been put. Overall space constraints, together with the judgement that readers coming to the existing literature from a philosophy of language background will find it harder to gain an overview of the different kinds of many-valued systems than to find information on applications of one or other of these systems to particular topics of interest, led to the decision to devote the bulk of the available space to §2.

2 Systems of Many-Valued Logic

2.1 Three Values

Suppose we take the language of classical propositional logic, and give it a semantics which countenances a third truth value—which we shall write as $*$—as well as the classical 1 and 0. Taking a (three-valued) model to be a free assignment of one of these three values to each basic proposition, the classical truth tables will no longer determine a truth value for every proposition of the language. For example, if p has the value $*$ on a given model, then the tables do not specify a value for compound propositions (e.g. $\neg p$, $p \rightarrow q$) which have p as a component. One way to remedy this situation is to replace the classical tables with three-valued truth tables. There are many such tables (see Bolc and Borowik [1992, ch.3] for a survey); we begin with the three that have played the most prominent role in the literature. Figure 5.6.4 shows the Bochvar (aka Kleene weak) tables [Rescher, 1969, 29–30] [Kleene, 1952, 334]. The idea here is that a compound whose components all take classical values takes the value that the classical tables assign, while if any of its components takes the value $*$, the compound takes the value $*$. Figure 5.6.5 shows the Kleene (strong) tables [Kleene, 1952, 334]. As before, a compound whose components all take classical values takes the value that the classical tables assign. As for the case where a component has the value $*$, there are two useful ways of thinking about what the tables dictate. First, we can think of the $*$ as an *unknown classical* value. So we suppose the $*$ is a 1 and calculate the value that the classical table would give, and we suppose the $*$ is a 0 and calculate the value that the classical table would give: if we get 1 both times, then that is the value in the new table; if we get 0 both times, then that is the value in the new table; and if we get 1 once and 0 once, then the value in

α	β	$\neg\alpha$	$\alpha \wedge \beta$	$\alpha \vee \beta$	$\alpha \rightarrow \beta$	$\alpha \leftrightarrow \beta$
1	1	0	1	1	1	1
1	*		*	*	*	*
1	0		0	1	0	0
*	1	*	*	*	*	*
*	*		*	*	*	*
*	0		*	*	*	*
0	1	1	0	1	1	0
0	*		*	*	*	*
0	0		0	0	1	1

Figure 5.6.4 Bochvar (aka Kleene weak) tables

α	β	$\neg\alpha$	$\alpha \wedge \beta$	$\alpha \vee \beta$	$\alpha \rightarrow \beta$	$\alpha \leftrightarrow \beta$
1	1	0	1	1	1	1
1	*		*	1	*	*
1	0		0	1	0	0
*	1	*	*	1	1	*
*	*		*	*	*	*
*	0		0	*	*	*
0	1	1	0	1	1	0
0	*		0	*	1	*
0	0		0	0	1	1

Figure 5.6.5 Kleene (strong) tables

the new table is *. (When we are trying to determine what value should be in the new table where *two* components both have the value *, we calculate all four possibilities given by replacing each * with a 1 or a 0.) Second, we can think of * as lying *between* 1 and 0 on a scale from more true to less true: * is more true than 0 but less true than 1. Then we can see the conjunction as taking the *least* true of the values of its conjuncts, the disjunction as taking the *most* true of the values of its conjuncts, and the conditional and biconditional as defined in the standard classical way from the other connectives: $\alpha \rightarrow \beta =_{df} \neg\alpha \vee \beta$ (or $\neg(\alpha \wedge \neg\beta)$) and $\alpha \leftrightarrow \beta =_{df} (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$. Figure 5.6.6 shows the Łukasiewicz tables [Łukasiewicz and Tarski, 1930]. These are exactly like the Kleene tables except for the values of $\alpha \rightarrow \beta$ and $\alpha \leftrightarrow \beta$ when α and β both have the value *: in the Łukasiewicz tables these values are 1, whereas in the Kleene tables they are *.

We come now to the third stage of presenting a semantics: dening notions of *tautology* and/or *consequence*. The standard way of defining these notions is to specify a subset of the truth values as *designated*. A tautology is then a proposition which takes a designated value on every model, and α is a logical consequence of Γ iff, on every model on which every proposition in Γ has a designated value, α has a designated value. (The

α	β	$\neg\alpha$	$\alpha \wedge \beta$	$\alpha \vee \beta$	$\alpha \rightarrow \beta$	$\alpha \leftrightarrow \beta$
1	1	0	1	1	1	1
1	*		*	1	*	*
1	0		0	1	0	0
*	1	*	*	1	1	*
*	*		*	*	1	1
*	0		0	*	*	*
0	1	1	0	1	1	0
0	*		0	*	1	*
0	0		0	0	1	1

Figure 5.6.6 Łukasiewicz tables

earlier definitions in the classical case emerge from this template by setting 1 as the only designated value among the two classical values 1 and 0. Marking designated values by underlining, we represent this choice as {0, 1}.) In the three-valued case, the most obvious choice is to set 1 as the only designated value: {0, ∗, 1}. However it is also not unreasonable to set both 1 and ∗ as designated values: {0, ∗, 1}. On the former choice, a tautology is a proposition which is *always true* (and consequence is a matter of preservation of truth); on the latter choice, a tautology is a proposition which is *never false* (and consequence is a matter of preservation of non-falsity). By combining different choices of designated values with different truth tables, we get different logics:

B_3: {0, ∗, 1} and Bochvar tables
B'_3: {0, ∗, 1} and Bochvar tables
K_3: {0, ∗, 1} and Kleene tables
K'_3, aka LP [Priest, 2008, 124]: {0, ∗, 1} and Kleene tables
$Ł_3$: {0, ∗ 1} and Łukasiewicz tables
$Ł'_3$: {0, ∗, 1} and Łukasiewicz tables

These logics differ more or less from classical logic. For a start, it is not hard to see that any semantics in which 0 is not a designated value and whose tables agree with the classical tables where only 1's and 0's are involved—this includes all the systems just introduced—will be such that all its tautologies are classical tautologies. Going the other way, all classical tautologies come out as tautologies of B'_3 and K'_3 (given the original stipulation that a many-valued logic is one which arises from a many-valued semantics and does not also arise from any two-valued semantics, this means that these logics—considered as sets of tautologies (the situation is different when we consider consequence relations)—are *not* many-valued logics: for while they do arise from many-valued semantics, they also arise from the classical two-valued semantics); B_3 and K_3, however, have no tautologies at all; $Ł_3$ has some of the classical tautologies (e.g. $p \rightarrow p$) but not all (e.g. $p \vee \neg p$ has the value ∗ when p does); $Ł'_3$ has all the tautologies of $Ł_3$ and some more besides (e.g. $p \vee \neg p$), but it still does not have all the classical tautologies (e.g. $\neg(p \rightarrow \neg p) \vee \neg(\neg p \rightarrow p)$ has the value 0 when p has the value ∗). The story is different again when it comes to consequence relations; for example, while B'_3 and K'_3 have the same tautologies as classical logic, they do not have classical consequence relations (e.g. q is a consequence of $\{p \wedge \neg p\}$ in classical logic, but not in B'_3 or K'_3: consider a model on which p has the value ∗—hence so does $p \wedge \neg p$—and q has the value 0).

2.2 Finitely Many Values

The three-valued tables of the previous section can all be generalized to the case where we have any finite number n of truth values. It is convenient to represent these n values as fractions:

$$\frac{0}{n-1}, \frac{1}{n-1}, \ldots, \frac{n-2}{n-1}, \frac{n-1}{n-1}$$

So where $n = 5$, for example, we have the following values: 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1. To generalize the Bochvar tables, we say that $[\alpha]$ is whatever the classical tables dictate when all α's components have the values 0 or 1; and otherwise it is the 'middle value'. (Where the number

n of values is odd, the middle value is simply $\frac{1}{2}$, i.e. half of the top value 1; where n is even, the 'middle' value is taken to be $\frac{n-2}{2(n-1)}$, i.e. half of the second-top value [Rescher, 1969, 43–4].) To generalize the Kleene tables, we use the rules in Figure 5.6.7. To generalize the Łukasiewicz tables, we use rules exactly like those for the Kleene systems, except that we replace the rule for the conditional with the following:

$$[\alpha \to \beta] = \min(1, 1 - [\alpha] + [\beta])$$

(In the rule for the biconditional in the Łukasiewicz systems, the conditional is then this Łukasiewicz conditional, not the Kleene conditional.) The idea here is that if the consequent is at least as true as the antecedent, then the conditional is completely true, while if the antecedent is truer than the consequent—and the difference between their truth values is k—then the conditional has the value $1 - k$, that is, it is k less than fully true. It is routine to verify that if we set $n = 3$, then the three sets of rules just given specify the three sets of tables given in the previous section (with $*$ written as $\frac{1}{2}$) and if we set $n = 2$, then these rules all specify the classical truth tables.

If we now choose a set of designated values, logics emerge. For example, taking the Łukasiewicz rules together with 1 as the only designated value yields the n-valued logics $Ł_n$. Lindenbaum showed that the set of tautologies of $Ł_m$ is a subset of the set of tautologies of $Ł_n$ $(m, n \geq 2)$ just in case $n - 1$ is a divisor of $m - 1$ [Łukasiewicz and Tarski, 1930, 48]. So, for example, the tautologies of $Ł_9$ are a subset of those of $Ł_5$ (and likewise of $Ł_3$), because 4 (and 2) is a divisor of 8. It follows that none of these many-valued logics has more tautologies than classical (two-valued) logic. The logics $Ł_n$ are all finitely axiomatizable (see e.g. Malinowski [1993, 39]), but not every finitely many-valued logic is; for example, Rescher [1969, 157–9] presents a three-valued logic which is not finitely axiomatizable (given substitution and modus ponens as rules of inference).

We turn now to a different strategy for generating n-valued systems, due to Post [1921]. Note that in the Kleene and Łukasiewicz (but not Bochvar) systems, the truth value of $\neg\alpha$ is as far below 1 as the truth value of α is above 0; that is, the distance between 0 and $[\alpha]$ is the same as the distance between $[\neg\alpha]$ and 1. This treatment of negation requires that there be a meaningful notion of *distance* between the truth values of the system. In Post's systems, by contrast, the truth values are merely *ordered*: given any two truth values, we can say which of them is the truer; but we cannot compare the *distances* between different pairs of values. We represent the truth values of the n-valued Post system as follows:

$$t_1, t_2, \ldots, t_n$$

$$\begin{array}{lcl}
[\neg\alpha] & = & 1 - [\alpha] \\
[\alpha \wedge \beta] & = & \min([\alpha], [\beta]) \\
[\alpha \vee \beta] & = & \max([\alpha], [\beta]) \\
[\alpha \to \beta] & = & [\neg\alpha \vee \beta] \\
[\alpha \leftrightarrow \beta] & = & [(\alpha \to \beta) \wedge (\beta \to \alpha)]
\end{array}$$

Figure 5.6.7 Rules for n-valued Kleene systems

The ordering of the values is this: t_i is less true than t_j just in case $i < j$. Figure 5.6.8 shows the truth functions of this system. So the value of $\neg\alpha$ is the value immediately after the value of α (in the ordering of the truth values from least true to most true), except in the case where the value of α is the top value t_n, in which case the value of $\neg\alpha$ is the bottom value t_1. The rule for disjunction is familiar from Kleene and Łukasiewicz: the value of the disjunction is the truer of the values of the disjuncts. The remaining connectives are defined in the standard ways from $\neg$ and $\vee$. Taking Post's rules and setting $n = 3$ yields a three-valued system different from any of the systems examined in §2.1; setting $n = 2$ yields classical logic.

It is a familiar fact that in classical logic, some connectives can be defined in terms of others. Similar kinds of results hold in many-valued logics. For example, Łukasiewicz took $\neg$ and $\rightarrow$ as primitive connectives, and defined the others in terms of them (see Figure 5.6.9). Similarly, Bochvar took $\neg$ and $\wedge$ as primitive, and Post took $\neg$ and $\vee$ as primitive. It is also sometimes possible to define the connectives of one system within another system. For example, Kleene conditional is definable in terms of Kleene negation and disjunction (Figure 5.6.7); but Kleene negation and disjunction coincide with Łukasiewicz negation and disjunction; so Kleene conditional is definable in terms of Łukasiewicz negation and disjunction. In fact, all the Kleene connectives are definable in terms of Łukasiewicz connectives but not vice versa; and all the Bochvar connectives are definable in terms of Kleene (and hence Łukasiewicz) connectives but not vice versa [Bergmann, 2008, 91–2]. A set of connectives is said to be *functionally complete* if we can define *all* possible connectives from the connectives in that set. In an n-valued system, an m-place connective has a truth table with n^m rows (one for each possible assignment of the n truth values to the m component propositions). We specify an m-place connective by putting a truth value in each row of the table. Thus there are $n^{(n^m)}$ possible m-place connectives. It is well known that in the classical (two-valued) case, all these connectives can be defined using only the five standard connectives introduced at the outset. (Indeed, we do not need all of them: just $\neg$ and $\wedge$ will do, or $\neg$ and $\rightarrow$, etc.) This property of functional completeness of (some subset of) the set of five standard connectives carries over to some, but not all, many-valued logics. In every n-valued Post

$$
\begin{array}{lcl}
\neg t_i & = & \begin{cases} t_{i+1} & \text{if } i \neq n \\ t_1 & \text{if } i = n \end{cases} \\
t_i \vee t_j & = & t_{\max(i,j)} \\
t_i \wedge t_j & = & \neg(\neg t_i \vee \neg t_j) \\
t_i \rightarrow t_j & = & \neg t_i \vee t_j \\
t_i \leftrightarrow t_j & = & (t_i \rightarrow t_j) \wedge (t_j \rightarrow t_i)
\end{array}
$$

Figure 5.6.8 Truth functions for n-valued Post systems

$$
\begin{array}{lcl}
\alpha \vee \beta & =_{df} & (\alpha \rightarrow \beta) \rightarrow \beta \\
\alpha \wedge \beta & =_{df} & \neg(\neg\alpha \vee \neg\beta) \\
\alpha \leftrightarrow \beta & =_{df} & (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)
\end{array}
$$

Figure 5.6.9 Defined connectives in Łukasiewicz systems

system, the set containing $\neg$ and $\vee$ is functionally complete. In no n-valued Łukasiewicz system is the set of five standard connectives functionally complete—although for some (but not all) n this set is precomplete, which means that it is not functionally complete but becomes so with the addition of *any* connective which is not already definable in terms of the standard five [Urquhart, 2001, 266–8].

In all the n-valued systems that we have considered so far, the set of truth values is linearly ordered—but it need not be so. For example, Belnap [1977] considers a system with four truth values—*T*, *F*, *B* and *N*—ordered as in Figure 5.6.10 (with $x < y$ iff one can get from x to y by following arrows). *N* and *B* are distinct truth values, with neither greater than the other—so we cannot take conjunction as min and disjunction as max. However, every pair of values has a supremum (a least value that is greater than or equal to both values in the pair) and an inmum (a greatest value that is less than or equal to both values in the pair). A partially ordered structure with this property is called a lattice. In any lattice of truth values, the strategy of defining conjunction and disjunction as min and max generalizes to defining them as inf and sup. So, for example, we can take $N \wedge B = F$ and $T \wedge B = B$, and $N \vee B = T$ and $N \vee F = N$.

We turn now to a new strategy for generating finitely many-valued systems, due to Jaśkowski [1936]. Suppose we have an m-valued system $\mathcal{X}$ and an n-valued system $\mathcal{Y}$. We can then form a new, $(m \times n)$-valued system $\mathcal{X} \times \mathcal{Y}$ by taking the *product* of these two systems. The truth values of the product system will be elements of the Cartesian product $\mathcal{X}_v \times \mathcal{Y}_v$, where $\mathcal{X}_v$ ($\mathcal{Y}_v$) is the set of values of $\mathcal{X}$ ($\mathcal{Y}$); that is, they will be pairs whose first element is a value of $\mathcal{X}$ and whose second element is a value of $\mathcal{Y}$. Truth functions of the product system are then specified coordinatewise (Figure 5.6.11: superscripts on truth functions indicate to which system they belong). One system of particular interest is the product of the classical system with itself. This system has four values:

$\langle 0, 0\rangle, \langle 0, 1\rangle, \langle 1, 0\rangle, \langle 1, 1\rangle$

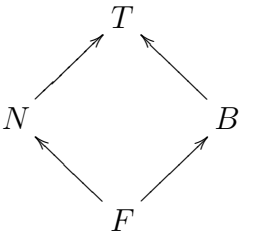


Figure 5.6.10 A lattice of four truth values

$$
\begin{array}{lcl}
\neg\langle x, y\rangle & = & \langle \overset{X}{\neg} x, \overset{Y}{\neg} y\rangle \\
\langle x_1, y_1\rangle \wedge \langle x_2, y_2\rangle & = & \langle x_1 \overset{X}{\wedge} x_2, y_1 \overset{Y}{\wedge} y_2\rangle \\
\langle x_1, y_1\rangle \vee \langle x_2, y_2\rangle & = & \langle x_1 \overset{X}{\vee} x_2, y_1 \overset{Y}{\vee} y_2\rangle \\
\langle x_1, y_1\rangle \rightarrow \langle x_2, y_2\rangle & = & \langle x_1 \overset{X}{\rightarrow} x_2, y_1 \overset{Y}{\rightarrow} y_2\rangle \\
\langle x_1, y_1\rangle \leftrightarrow \langle x_2, y_2\rangle & = & \langle x_1 \overset{X}{\leftrightarrow} x_2, y_1 \overset{Y}{\leftrightarrow} y_2\rangle
\end{array}
$$

Figure 5.6.11 Truth functions for product systems

Its truth table for negation (for example) is shown in Figure 5.6.12. Note that its truth tables for conjunction and disjunction could alternatively be arrived at by setting $\langle x_1, y_1\rangle \leq \langle x_2, y_2\rangle$ iff ($x_1 \leq x_2$ and $y_1 \leq y_2$), and then following the inf/sup strategy of the previous paragraph. When it comes to choosing designated values for a product system, there are two obvious choices: $\langle x, y\rangle$ is designated iff x is designated in system $\mathbf{X}$ and/or (choose one) y is designated in system $\mathbf{Y}$. In a *self* -product system $\mathbf{X} \times \mathbf{X}$, both choices yield the same set of tautologies—which is simply the set of tautologies of $\mathbf{X}$ [Rescher, 1969, 101].

2.3 Infinitely Many Values

If we wish to countenance more than finitely many truth values, there are two particularly natural options: we can take as truth values all the *rational* numbers between 0 and 1 inclusive, or the *real* interval [0, 1] (containing all the real numbers between 0 and 1 inclusive). Łukasiewicz considered both these options. The rules for assigning truth values to compound propositions stated in §2.2 carry over unchanged. If we take 1 as the only designated values, the two resulting logics are named $Ł_{\aleph_0}$ and $Ł_c$ respectively. Interestingly, they both have the same set of tautologies [Rescher, 1969, 38–9].

A semantics which takes [0, 1] as its set of truth values (and treats the connectives truth-functionally) is called a fuzzy semantics; a logic that arises from a fuzzy semantics is a fuzzy logic. Among philosophers, the best-known rules for assigning truth values to compound propositions in a fuzzy semantics are the Zadeh rules; the statement of these rules is exactly the same as that of the Kleene rules in Figure 5.6.7. Among logicians, the fuzzy systems of most interest are the t-norm fuzzy logics. Where $\overset{\star}{\wedge}$ is a binary function on [0, 1], which we are going to use to define conjunction, it is natural to want $\overset{\star}{\wedge}$ to satisfy the conditions shown in Figure 5.6.13. A binary function on [0, 1] which satises these conditions

α	$\neg\alpha$
$\langle 0,0\rangle$	$\langle 1,1\rangle$
$\langle 0,1\rangle$	$\langle 1,0\rangle$
$\langle 1,0\rangle$	$\langle 0,1\rangle$
$\langle 1,1\rangle$	$\langle 0,0\rangle$

Figure 5.6.12 Negation in the self-product of classical logic

$$
\begin{array}{rcl}
x \overset{\star}{\wedge} y & = & y \overset{\star}{\wedge} x \\
(x \overset{\star}{\wedge} y) \overset{\star}{\wedge} z & = & x \overset{\star}{\wedge} (y \overset{\star}{\wedge} z) \\
x_1 \leq x_2 & \text{implies} & x_1 \overset{\star}{\wedge} y \leq x_2 \overset{\star}{\wedge} y \\
y_1 \leq y_2 & \text{implies} & x \overset{\star}{\wedge} y_1 \leq x \overset{\star}{\wedge} y_2 \\
1 \overset{\star}{\wedge} x & = & x \\
0 \overset{\star}{\wedge} x & = & 0
\end{array}
$$

Figure 5.6.13 Conditions on t-norms

is called a *t-norm*. If it is also continuous—in the usual sense—then it is a continuous t-norm. A t-norm fuzzy logic is determined by a choice of a continuous t-norm as the truth function for conjunction: truth functions for the other connectives are then defined in terms of this t-norm. The truth function $\overset{\star}{\rightarrow}$ for conditional is the *residuum* of the t-norm:

$$x \overset{\star}{\rightarrow} y = \max\{z\colon x \overset{\star}{\wedge} z \leq y\}$$

and the truth function $\overset{\star}{\neg}$ for negation is the *precomplement* of this conditional:

$$\overset{\star}{\neg} x = x \overset{\star}{\rightarrow} 0$$

Biconditional is then defined in the standard way using conditional and conjunction. Disjunction will be discussed below. The most important examples of continuous t-norms are shown in Figure 5.6.14. These three are fundamental in the sense that every continuous t-norm is a combination of them [Hájek, 1998, 32]. Note that the Gödel t-norm is the min operation used to define conjunction in Zadeh/Kleene logic; it is the only *idempotent* t-norm (i.e. one which satisfies the condition $x \overset{\star}{\wedge} x = x$) [Klir and Yuan, 1995, 63]. The conditionals and negations derived from these three t-norms (as residuum and precomplement) are shown in Figure 5.6.15. The Łukasiewicz operations are the familiar ones—which is why the t-norm that gives rise to them is called the Łukasiewicz t-norm. Note however that this t-norm is *not* the min operation used to define conjunction in Łukasiewicz's many-valued logics (§2.2): that min operation is the Gödel t-norm, and the conditional and negation to which it gives rise are not Łukasiewicz's. The Gödel operations are named for their discussion in Gödel [1986]. Note that while the Gödel conjunction is the same as the conjunction in Zadeh/Kleene logic, the conditionals and negations in these logics are different. The conditional arising from the product t-norm was discussed in Goguen [1968–69] and is often referred to as the Goguen conditional.

$$\begin{array}{ll}
\text{Łukasiewicz:} & x \overset{\star}{\wedge} y = \max(0, x + y - 1) \\
\text{Gödel:} & x \overset{\star}{\wedge} y = \min(x, y) \\
\text{Product:} & x \overset{\star}{\wedge} y = x \cdot y
\end{array}$$

Figure 5.6.14 Important t-norms

$$\begin{array}{lll}
\text{Łukasiewicz:} & x \overset{\star}{\rightarrow} y = \begin{cases} 1 & \text{if } x \leq y \\ 1 - x + y & \text{if } x > y \end{cases} & \overset{\star}{\neg} x = 1 - x \\
\text{Gödel:} & x \overset{\star}{\rightarrow} y = \begin{cases} 1 & \text{if } x \leq y \\ y & \text{if } x > y \end{cases} & \overset{\star}{\neg} x = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \\
\text{Product:} & x \overset{\star}{\rightarrow} y = \begin{cases} 1 & \text{if } x \leq y \\ y/x & \text{if } x > y \end{cases} & \overset{\star}{\neg} x = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}
\end{array}$$

Figure 5.6.15 Residua and precomplements

It is common to add an additional 'weak' conjunction to each of these fuzzy logics—with the t-norm conjunction then being termed 'strong'—defined as follows (with subscripts *w* and *s* indicating strong and weak):

$$(\alpha \wedge_w \beta) =_{df} \alpha \wedge_s (\alpha \to \beta)$$

Whatever continuous t-norm is used for $\wedge_s$ here (and where $\to$ is its residuum), it turns out that *weak* conjunction is always the min operation [Hájek, 1998, 36]. Disjunction also comes in strong and weak forms. Weak disjunction is defined in terms of weak conjunction and conditional:

$$(\alpha \vee_w \beta) =_{df} ((\alpha \to \beta) \to \beta) \wedge_w ((\beta \to \alpha) \to \alpha)$$

It always turns out to be the max operation [ibid.]. Strong disjunction is defined as the dual of strong conjunction:

$$(x \vee_s y) = 1 - ((1 - x) \wedge_s (1 - y))$$

(Where the strong conjunction is a t-norm, this dual operation will always be a t-conorm. A t-conorm is a binary operation $\overset{\star}{\vee}$ on [0, 1] which satisfies conditions exactly like those for a t-norm—i.e. put $\overset{\star}{\vee}$ for $\overset{\star}{\wedge}$ throughout Figure 5.6.13—*except* that we replace the final two conditions with $1 \overset{\star}{\vee} x = 1$ and $0 \overset{\star}{\vee} x = x$.) In our three systems, this definition yields the operations shown in Figure 5.6.16. The duality of strong conjunction and strong disjunction means that where negation is defined as $\neg x = 1 - x$—for example, in Łukasiewicz t-norm logic—the de Morgan laws will hold (when stated in terms of negation, strong conjunction and strong disjunction).

Hájek [1998] introduces an axiom system for a logic BL (basic logic) which proves all and only those formulas which come out as tautologies (taking 1 as the only designated value) no matter what continuous t-norm is taken as the truth function for (strong) conjunction. Each of the three specific t-norm logics mentioned above is then axiomatized by the addition of an axiom or axioms—different in each case—to BL.

Thus far we have mentioned only propositional logics. Most many-valued semantics for propositional logic can be generalized to predicate logic by following through the following fundamental idea. We identify a subset *S* of a background set *U* with its characteristic function f_S: a function from *U* to the set of truth values. Where *x* is an object in *U* and *y* is a truth value, $f_S(x) = y$ means that *x* is a member of *S* to degree *y*. So the truth values now function as values or degrees of *membership* (of objects in sets), as well as values or degrees of *truth* (of propositions). An *n*-place predicate is assigned an *extension* as its semantic value: a subset of the set of *n*-tuples of members of the domain; that is, a function from this set of *n*-tuples to the set of truth values. The truth value of

Łukasiewicz:	$x \vee_s y = \min(1, x + y)$
Gödel:	$x \vee_s y = \max(x, y)$
Product:	$x \vee_s y = x + y - x \cdot y$

Figure 5.6.16 t-conorms/strong disjunctions

the atomic formula *Rab*, for example—comprising a two-place predicate *R* followed by two names *a* and *b*—is then whatever value the extension of *R* assigns to the ordered pair comprising the referent of *a* followed by the referent of *b*. In order to give rules for assigning truth values to universally and existentially quantied formulas, we need to generalize the conjunction and disjunction operations (respectively) so that they assign values to sets—not just pairs—of values. For details, see Smith [2008, §1.2, §2.2].

2.4 Non-Truth-Functional Systems

All the systems we have looked at so far are truth-functional: the truth value of a compound proposition depends only on the truth values of its components; that is, the semantics of the connectives are given by associating them with truth functions. One notable method for assigning multiple values to propositions in a *non*-truth-functional way is that of probability logic, where values—probabilities—are assigned in accordance with the probability calculus. The value of $p \vee q$, for example, does not then depend only on the values of p and q: it also depends on the content of p and q, and in particular on whether these propositions are mutually exclusive. Interestingly, the set of tautologies of probability logic (with 1 as the only designated value) is the same as the set of tautologies of classical logic [Rescher, 1969, 186–7]. Another notable method is that of supervaluations [van Fraassen, 1966]. We begin with a three-valued model (i.e. a mapping from basic propositions to the values 0, 1 and $*$). Say that a classical model M_2 extends a three-valued model M_3 iff M_2 is exactly like M_3 except that where M_3 assigns $*$ to a basic proposition, M_2 assigns 1 or 0. The rule for assigning truth values to compound propositions is then this: $[\alpha] = 1$ (0) on M_3 iff $[\alpha] = 1$ (0) on *every* classical model which extends M_3. So, for example, where p and q both have the value $*$, $p \wedge q$ will also have the value $*$; yet even though $\neg p$ will have the same value as q, $p \wedge \neg p$ will have the value 0. So the system is not truth-functional. A variant of supervaluationism is subvaluationism, which differs by having 'some' in place of 'every' in the rule for assigning truth values to compound propositions: $[\alpha] = 1$ (0) on M_3 iff $[\alpha] = 1$ (0) on *some* classical model which extends M_3. If we take 1 as the only designated value, then the (single-conclusion) consequence relation arising from the supervaluationist semantics is identical to the classical consequence relation (see e.g. Smith [2008, 82]), while that arising from the subvaluationist semantics is not (e.g. adjunction fails: $\{\alpha, \beta\} \not\models \alpha \wedge \beta$). However, if we move to a multiple-conclusion consequence relation, then the duality between supervaluationism and subvaluationism is restored: neither yields the classical consequence relation [Hyde, 1997]. Another variant is the degree-theoretic form of supervaluationism, which assigns values in the set [0, 1]. We start with a fuzzy model (i.e. a mapping from basic propositions to [0, 1]), introduce a measure on the set of classical extensions of this fuzzy model, and then say that the degree of truth of a compound sentence on the fuzzy model is equal to the measure of the set of classical extensions on which it is true. For further details see Smith [2008, §2.4.1].

2.5 Many-Valued Systems in the Loose Sense

Some of the motivations for many-valued systems to be discussed in §3 lead most naturally not to the idea of additional truth values, alongside the classical 0 and 1, but to the idea of propositions possessing *neither* (or *both*) of the classical values. This idea leads to what we termed at the outset many-valued systems in the loose sense. Instead of adding

truth values, we stick with the classical set {0, 1}, but we alter the definition of a model. Instead of consisting in a *function* which assigns a truth value to each basic proposition, it now consists in a *relation* between the set of basic propositions and the set of truth values. Where a proposition is related to just *one* truth value we have the analogues of classical truth and falsity; where a proposition is related to *neither* truth value we have a truth value *gap*; where a proposition is related to *both* truth values we have a truth value *glut*. Allowing gaps or gluts but not both yields a three-status system; allowing both yields a four-status system. The question of how to assign truth values to compound propositions when some of their components have gaps or gluts may be addressed in ways analogous to any of those already considered in relation to three- or four-valued systems: no essentially new ideas are involved—simply the translation of existing ideas into the new setting (e.g. we might employ truth tables—with empty and/or multiply-filled cells in place of third and/or fourth values—or we might employ a version of supervaluationism; for further details see e.g. Smith [2008, §2.3, §2.4]). Some writers (e.g. Kripke [1975, 700, n.18]; cf. also van Fraassen [1974, 231]) are adamant that there is a deep *conceptual* difference between positing a third truth value and allowing truth value gaps, but in practice it is often more straightforward to formulate systems with additional values and assignment functions rather than two values and assignment relations, and so even when the guiding idea is many-valued in the loose sense, the implementation is often many-valued in the strong sense (e.g. Blamey [1986], Langholm [1988]).

3 Uses of Many-Valued Logics

One of the earliest motivations for many-valued logics—going back to Aristotle [Łukasiewicz, 1930, 63–4]—was the consideration of modal propositions (e.g. 'It is possible that *p*') and propositions concerning contingent future events (e.g. 'I shall be in Warsaw at noon tomorrow'). In relation to the former, Łukasiewicz [1930, 51] saw that the semantics of the operator 'It is possible that' cannot plausibly be given by any of the four possible one-place two-valued truth functions; in relation to the latter, he argued that assigning 'I shall be in Warsaw at noon tomorrow' either 1 or 0 as truth value now conflicts with the fact that it is possible, but not necessary, that I shall be in Warsaw at noon tomorrow, and he writes that his "three-valued system of propositional logic owes its origin to this line of thought" [53]. This was also an important motivation for Peirce and MacColl (see Rescher [1969, 4–5] for references).

The consideration of probabilistic propositions, and in some cases propositions arising from quantum mechanics, was a motivation for MacColl, Zawirski, Reichenbach, and Carnap (see Rescher [1969, 14–5, 184–8, 210–1] and Malinowski [1993, 66] for references).

Consideration of the paradoxes—set-theoretic (e.g. Russell's) and/or semantic (e.g. the Liar, where it seems impossible to assign either truth value 1 or 0 to 'This sentence is false')—was a motivation for Bochvar, Moh Shaw- Kwei and others (see Rescher [1969, 13, 29, 207] for references).

Kleene was motivated by a consideration of partial recursive predicates. Think of such a predicate as coming with an algorithm. Given some objects as input, the algorithm terminates with the answer 'Yes': the predicate is true of these objects. Given some (other) objects as input, the algorithm terminates with the answer 'No': the predicate is false of these objects. But for some (other) objects, the algorithm does not terminate with either answer: the predicate is *undefined* for these objects. Hence Kleene's three

truth values: 𝔱 (true), 𝔣 (false) and 𝔲 (undefined). A related application of many-valued logics is in the analysis of category mistakes arising from predicates which are truly applicable to some objects, falsely applicable to others, and *inapplicable* to yet other objects [Rescher, 1969, 28].

Kleene [1952, 335] also considers a different interpretation of his three values: "𝔱, 𝔣, 𝔲 must be susceptible of another meaning besides (i) 'true', 'false', 'undefined', namely (ii) 'true', 'false', 'unknown (or value immaterial)'. Here 'unknown' is a category into which we can regard any proposition as falling, whose value we either do not know or choose for the moment to disregard; and it does not then exclude the other two possibilities 'true' and 'false'." This kind of *epistemic* application of many-valued logics is also found, for example, in Belnap [1977].

Many-valued logics have important applications in proof theory. For example, they can be used to show that certain axioms in a given axiomatization (e.g. of *classical* propositional logic) are *independent* of others. This application originates with Bernays; for illustrative examples see Malinowski [1993, 105–6] and Mendelson [1997, 43–5] (and for some examples from modal logic see Lewis and Langford [1932, Appendix II]). They can also be used for showing the *consistency* of a set of axioms (i.e. that not everything can be deduced from it). This application originates with Post [Rescher, 1969, 106].

Like classical logic, many-valued logics have applications in circuit/switching theory; for discussion and references see Malinowski [1993, 109–10] and Hähnle [2001, §8]. They have also been applied to scheduling problems [Ackermann, 1967, 78]. Many-valued logics have been employed in automated theorem-proving; for a discussion of the many further applications of many-valued theorem provers see Hähnle [1993, ch.7].

Many-valued logics have been applied to the analysis of presupposition. 'Bill knows that Ben is late' presupposes that Ben is late; 'The king of France is bald' presupposes that there is a king of France; and so on. The many-valued approach to presupposition holds that if one of a sentence's presuppositions is false, then the sentence is neither true nor false. This is known as a *semantic* approach to presupposition, because it holds that presupposition failure affects the *truth value* of a sentence. For further details see Gamut [1991, 178–90, 212–4].

A closely related use of many-valued logics is in the treatment of singular terms which lack referents (e.g. names such as 'Pegasus' and definite descriptions such as 'the king of France'). On this approach, atomic statements involving such terms are neither true nor false. This idea, which can be found for example in Strawson [1950], was the motivation for the development of supervaluationism by van Fraassen [1966].

The final use of many-valued logics which we shall mention here is in the treatment of vagueness, where the basic thought is that statements about borderline cases of vague predicates (e.g. 'Bob is tall', where Bob is neither clearly tall nor clearly not tall) are neither true nor false. Supervaluationist and fuzzy approaches in particular have played an important role in the vagueness literature (see Smith [2008] for further discussion and references).

Acknowledgements

Thanks to Allen Hazen, Libor Běhounek, and Richard Zach for very helpful comments on an earlier draft. Thanks to the Australian Research Council for research support.

Related Topics

1.4 Presupposition
1.5 Implicature
1.8 Compositionality
1.12 Vagueness
1.13 Empty Names
2.1 Reference
5.8 Intuitionism.

References

Robert Ackermann. *An Introduction to Many-Valued Logics*. Routledge & Kegan Paul, London, 1967.
Nuel D. Belnap. A useful four-valued logic, 1977. In Dunn and Epstein [1977], pages 8–37.
Merrie Bergmann. *An Introduction to Many-Valued and Fuzzy Logic: Semantics, Algebras, and Derivation Systems*. Cambridge University Press, Cambridge, 2008.
Stephen Blamey. Partial logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III, pages 1–70. D. Reidel, Dordrecht, 1986.
Leonard Bolc and Piotr Borowik. *Many-Valued Logics*, volume 1. Springer- Verlag, Berlin, 1992.
J. Michael Dunn and George Epstein, editors. *Modern Uses of Multiple- Valued Logic*. D. Reidel, Dordrecht, 1977.
D.M. Gabbay and F. Guenthner, editors. *Handbook of Philosophical Logic*, volume 2. Kluwer, Dordrecht, 2nd edition, 2001.
L.T.F. Gamut. *Logic, Language, and Meaning*, volume I. University of Chicago Press, Chicago, 1991.
Kurt Gödel. On the intuitionistic propositional calculus. In Solomon Feferman et al., editor, *Kurt Gödel: Collected Works*, volume I, pages 222–5. Oxford University Press, New York, 1986. Originally published in 1932.
Joseph Goguen. The logic of inexact concepts. *Synthese*, 19:325–73, 1968–69.
Siegfried Gottwald. *A Treatise on Many-Valued Logics*. Research Studies Press, Baldock, 2001.
Reiner Hähnle. *Automated Deduction in Multiple-valued Logics*. Clarendon Press, Oxford, 1993.
Reiner Hähnle. Advanced many-valued logics, 2001. In Gabbay and Guenthner [2001], pages 297–395.
Petr Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.
Dominic Hyde. From heaps and gaps to heaps of gluts. *Mind*, 108:641–60, 1997.
Stanisław Jaśkowski. Investigations into the system of intuitionist logic, 1936. In McCall [1967], pages 259–63.
Stephen Cole Kleene. *Introduction to Metamathematics*. D. Van Nostrand, Princeton, 1952.
George J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, NJ, 1995.
Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.
Tore Langholm. *Partiality, Truth and Persistence*. CSLI, Stanford, 1988.
Clarence Irving Lewis and Cooper Harold Langford. *Symbolic Logic*. Century, New York, 1932.
Jan Łukasiewicz. Philosophical remarks on many-valued systems of propositional logic, 1930. In McCall [1967], pages 40–65.
Jan Łukasiewicz and Alfred Tarski. Investigations into the sentential calculus, 1930. In Tarski [1956], pages 38–59.
Grzegorz Malinowski. *Many-Valued Logics*. Clarendon Press, Oxford, 1993.
Storrs McCall, editor. *Polish Logic: 1920–1939*. Clarendon Press, Oxford, 1967.
Elliott Mendelson. *Introduction to Mathematical Logic*. Chapman & Hall, London, fourth edition, 1997.
George Metcalfe, Nicola Olivetti, and Dov Gabbay. *Proof Theory for Fuzzy Logics*. Springer, 2009.
Emil L. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163–85, 1921.
Graham Priest. *An Introduction to Non-Classical Logic: From Ifs to Is*. Cambridge University Press, Cambridge, second edition, 2008.
Nicholas Rescher. *Many-valued Logic*. McGraw-Hill, New York, 1969.
Nicholas J.J. Smith. *Vagueness and Degrees of Truth*. Oxford University Press, 2008.
P.F. Strawson. On referring. *Mind*, 59:320–44, 1950.

Alfred Tarski. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Clarendon Press, Oxford, 1956. Translated by J.H. Woodger.

Alasdair Urquhart. Basic many-valued logic, 2001. In Gabbay and Guenthner [2001], pages 249–95.

Bas C. van Fraassen. Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 63:481–95, 1966.

Bas C. van Fraassen. The labyrinth of quantum logics. In Robert S. Cohen and Marx W. Wartofsky, editors, *Logical and Epistemological Studies in Contemporary Physics* (Boston Studies in the Philosophy of Science, volume XIII), pages 224–54. D. Reidel, Dordrecht, 1974.

Robert G. Wolf. A survey of many-valued logic (1966–1974), 1977. In Dunn and Epstein [1977], pages 167–323.

Further Reading

For readers whose only background in logic is a typical introductory course or book: Bergmann [2008] provides a thorough introduction to Bochvar, Kleene and Łukasiewicz three-valued logics, and fuzzy logics.

For readers with greater background in logic: Malinowski [1993], Urquhart [2001] plus Hähnle [2001], and Gottwald [2001] provide wide-ranging surveys of the field, including important details about consequence relations, functional completeness, proof theory, decidability and complexity, and other topics of central importance in logic; for more on fuzzy logics see Hájek [1998] and Metcalfe et al. [2009].

For readers interested in the history of the subject: Rescher [1969] includes a comprehensive bibliography of work on many-valued logics to 1965; Wolf [1977] extends this to 1974.

For readers wishing to explore applications of many-valued logics with particular relevance to philosophy of language: Smith [2008] gives an extended argument for the conclusion that the correct account of vagueness must involve a many-valued semantics.

5.7
DYNAMIC LOGIC IN NATURAL LANGUAGE

Johan van Benthem

1 Dynamic Perspectives in Natural Language

Truth-conditional Semantics

Standard first-order logic defines truth in a three-part scheme: a language, structures $\boldsymbol{D}$ of objects with relations and operations, and maps from language to structures that drive semantic evaluation. In particular, "interpretation functions" I map predicate letters to real predicates, while variable assignments s map individual variables to objects. Logicians often lump $\boldsymbol{D}$ and I together into a "model" $\mathbf{M}$, and then interpret formulas:

formula φ *is true in model* $\mathbf{M}$ *under assignment* s ($\mathbf{M}, s \models \varphi$)

with a recursive definition matching syntactic construction steps with semantic operations for connectives and quantifiers. This pattern has been applied to natural language since Montague 1974, stating under which conditions a sentence is true. Compositional interpretation in tandem with syntactic construction works even beyond logical and natural languages: it is also a well-known design principle for programs (van Leeuwen 1990). And the paradigm finds an elegant mathematical expression in universal algebra and category theory.

From Products to Activities

Still, the above semantics merely describes a static relationship between sentences and the world. But truth is just one aspect of natural language, perhaps not even its crux. What makes language constitutive of human life seem *dynamic acts* of assertion, interpretation, or communication. In recent years, such acts have entered logical theory, from interpretation to speech acts and discourse. This is often considered pragmatics—but often, the natural semantic meaning *is* the use. In a maxim from mathematics and computer science: 'never study representation without transformation'. We cannot understand the static structure of a language without studying the major processes it is used for. Our very lexicon suggests a duality between product and process views. "Dance" is both a verb and a noun, "argument" is an activity one can pursue and its product that logicians write down.

Concrete Dynamic Systems

There are many sources for current dynamic semantics of natural language. One is the seminal work of Kamp and Heim on anaphoric interpretation of pronouns as creating *discourse representation structures* that get modified as speech proceeds, and that can be matched against reality when the need arises. But the most incisive example has been "dynamification" of existing logics. First-order logic is a pilot for static truth conditions, but it can also model essentials of the process of evaluation. The latter involves shifting relationships between variables and objects (Groenendijk and Stokhof 1991). Consider the truth condition for an existential quantifier:

$\mathbf{M}, s \models \exists x\, \varphi$ iff there is an object d in $\mathbf{M}$ such that $\mathbf{M}, s[x := d] \models \varphi$.

Intuitively, this calls for a search through available objects d in $\mathbf{M}$ for the variable x until we hit the first d for which φ holds. The latter object is then available for further reference, as it should, say, in little texts like "$\exists x Px.\ Qx$" ("A *man* came in. *He* whistled.") that support anaphoric reference between the two occurrences of x across sentence boundaries.

Dynamic ideas work much more widely, with temporal expressions as a major paradigm (ter Meulen 1995). As for sentence-level processes, sentences can change information states of hearers by elimination of all models from a current set that do not satisfy the formula. This folklore idea of range-with-elimination underlies the account of conversation in Stalnaker 1978, or the "update semantics" of Veltman 1996, where meanings are potentials for changing information states. Thus, dynamic linguistic acts come into the scope of logic, making them amenable to the compositional analysis that has served truth-conditional views in the past.

Richer Versions: Social Dynamics and Games

Classical semantics has no actors, since it is about bare relationships with reality. Dynamic semantics is about single agents that compute on discourse structures, or change single minds. But language is about speakers and hearers that create shared meanings, and over that channel, engage in meaningful activities, cooperative or competitive. This requires a study of information flow in multi-agent communication where "social" knowledge about what others know and mutual expectations are crucial. And beyond single speech acts, there is a longer-term strategic aspect. I choose my words toward an end, depending on how I think you will take them, and next, so do you. Such behavior over time is the realm of game theory. Dynamic logics and games will be discussed below.

Literature

For dynamic semantics, see Dekker 2008, Groenendijk et al. 1996, and for discourse representation theory, Kamp and Reyle 1993. Other sources include speech act theory (Searle and Vanderveken 1985) and "score-keeping games" (Lewis 1979). The "Dynamics" chapter in van Benthem and ter Meulen 1997 (updated in 2010 with an appendix on "Information Dynamics") adds links to dynamic logics, artificial intelligence, and computational linguistics. Kamp and Stokhof 2008 has extensive philosophical reflection.

2 Parallels with Computer Languages and Computation

The Unity of Languages

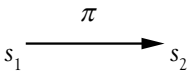
Philosophy has long known a tension between formal and natural language methods, with Russell's "Misleading Form Thesis" claiming that natural language obscures the logical form of statements. By contrast, Montague proclaimed the unity of formal and natural languages in their design principles and theoretical properties. Later authors tested "Montague's Thesis" on *programming languages*: formally designed, but driving a real communicative practice between humans and machines. They found striking analogies with natural language, from category structure to paradoxes of intensionality.

Semantics of Programs: Computing Change

A major challenge in logics of computation is giving a meaning to imperative programs. These are not propositions that are true or false but instructions for changing states of a computer, or indeed any process. As it turns out, first-order assignment semantics is a great fit here, allowing for a compositional definition of the relation of *successful transition* between assignments for program execution:

$s_1 [[\pi]]^{\mathbf{M}} s_2$: there is a successful execution of program π starting in s_1 and ending in s_2.

Here the assignments s, originally an auxiliary device, become important in their own right, as memory states of a computer. A typical case are atomic programs $x:=t$, where the assignment s_1 changes to a new s with all values for variables the same, except that x is now set to $[[t]]^{\mathbf{M}}{}_{s1}$, the value of the term t in $\mathbf{M}$ under the old assignment s_1. In a picture, we now view meanings in terms of transition arrows between states:



Thus, action is identified with pairs <*input state, output state*>. There are much richer process views in computer science, but we will use this simple extensional format in what follows.

Compositional structure

Like propositions, programs have complex syntactic structure, and their interpretation proceeds inductively: we match them with semantic ones. Here are three basic operations (the textbook Harel 1987 explains them with kitchen recipes):

Sequential composition	$\pi_1; \pi_2$
Guarded choice	*IF* φ *THEN* π_1 *ELSE* π_2
Iteration	*WHILE* φ *DO* π

For instance, the semi-colon ; denotes sequential composition of relations. Its transitions arise from first making a successful transition for π_1, and then one for π_2. The *WHILE* loop is unbounded; we make a computer run for as long as it takes to achieve $\neg\varphi$.

Logics of Programs

The oldest computational program logic is *Hoare calculus* of *correctness assertions* that express what a given program does in terms of standard propositions:

$\{\varphi\}\ \pi\ \{\psi\}$ "after *every* successful execution of program π starting from a state where *precondition* φ holds, *postcondition* ψ holds":

Note the co-existence of programs π and propositions φ: there is no conflict between statics and dynamics. Also, non-determinism is allowed: a program may have several executions. Now we can do program logic, much as rules for connectives analyze static propositions:

Example Rules of the Hoare Calculus.

$$\frac{\{\varphi\}\ \pi_1\ \{\psi\} \qquad \{\psi\}\ \pi_2\ \{\chi\}}{\{\varphi\}\ \pi_1\ ;\ \pi_2\ \{\chi\}} \quad \text{composition}$$

$$\frac{\{\varphi \wedge \chi\}\ \pi_1\ \{\psi\} \qquad \{\varphi \wedge \neg\chi\}\ \pi_2\ \{\psi\}}{\{\varphi\}\ \mathit{IF}\ \chi\ \mathit{THEN}\ \pi_1\ \mathit{ELSE}\ \pi_2\ \{\psi\}} \quad \text{guarded choice}$$

$$\frac{\{\varphi\}\ \pi\ \{\varphi\}}{\{\varphi\}\ \mathit{WHILE}\ \psi\ \mathit{DO}\ \pi\ \{\varphi \wedge \neg\psi\}} \quad \text{iteration}$$

Checking soundness of these inference rules will make you understand a lot about the logic of change.

From Programs to General Actions

The preceding describes a logic of action for any dynamic event, not just shifts in variable assignments or states of a computer. Logics like this have been applied to natural language, as we shall see, but also to conversation, strategies in games, and quantum-mechanical measurements that change a physical system. In this process, a reversal of perspective has occurred. Thinking of all the dynamic effects of using language, we might consider the computational stance as primary. Van Benthem 1996 even claims that *natural language is a programming language* for cognitive actions.

3 Technical Background: Dynamic Logic of Action

Behind program semantics lies a logic familiar to philosophers studying intensional notions.

Modal Logics of Process Graphs

Consider *process graphs* $\mathbf{M} = (S, \{R_a\}, V)$ with a set of states S, a family of binary transition relations R_a for basic actions (sometimes written $\rightarrow_a$), and a valuation V interpreting proposition letters p as local properties of states. Over such models, one can interpret a language with labeled modalities over "action-accessible states":

$\mathbf{M}, s \models [a]\varphi$ iff $\mathbf{M}, t \models \varphi$ for all t with $s\ R_a\ t$.

The dual existential modality $\langle a \rangle \varphi$ is defined as $\neg[a]\neg\varphi$. Hoare correctness statements are modal implications $\varphi \rightarrow [\pi]\psi$. (A good textbook is Blackburn et al. 2000.) Major uses of modal logic today are *action* and *knowledge*, notions that will return.

Complex Actions: Propositional Dynamic Logic

The same formalism can deal with complex actions. The language now has components, one of *programs (P)* and one of *formulas (F)*:

F := *atomic propositions* | $\neg F$ | $(F \wedge F)$ | $[P]F$
P := *atomic actions* | $(P;P)$ | $(P \cup P)$ | P^* | $F?$

For elegance and sweep, program operators are now the *regular operations* of composition, Boolean choice, Kleene iteration, and tests for formulas. The semantics matches the mutual recursion in the syntax. $\mathbf{M}, s \models \phi$ says that ϕ *is true at state* s, while $\mathbf{M}, s_1, s_2 \models \pi$ says the transition from s_1 to s_2 is a *successful execution* of program π. Here are a few key clauses:

- $\mathbf{M}, s \models [\pi]\phi$ iff *for all* s' *with* $\mathbf{M}, s, s' \models \pi$, *we have* $\mathbf{M}, s' \models \phi$

- $\mathbf{M}, s_1, s_2 \models a$ iff $(s_1, s_2) \in R_a$
  $\mathbf{M}, s_1, s_2 \models \pi_1;\pi_2$ iff *there is an* s_3 *with* $\mathbf{M}, s_1, s_3 \models \pi_1$ *and* $\mathbf{M}, s_3, s_2 \models \pi_2$
  $\mathbf{M}, s_1, s_2 \models \pi_1 \cup \pi_2$ iff $\mathbf{M}, s_1, s_2 \models \pi_1$ *or* $\mathbf{M}, s_1, s_2 \models \pi_2$
  $\mathbf{M}, s_1, s_2 \models \pi_1^*$ iff *some finite sequence of* π*–transitions in* $\mathbf{M}$ *connects the state* s_1 *with the state* s_2
  $\mathbf{M}, s_1, s_2 \models \phi?$ iff $s_1 = s_2$ *and* $\mathbf{M}, s_1 \models \phi$

Axiomatic System

Propositional dynamic logic has a natural proof system *PDL*. We give it here to show how logics of action and change can be designed just like classical logics:

- All principles of the minimal modal logic for all modalities $[\pi]$
- Computation rules for decomposing program structure:

$[\pi_1;\pi_2]\phi \leftrightarrow [\pi_1][\pi_2]\phi$
$[\pi_1 \cup \pi_2]\phi \leftrightarrow [\pi_1]\phi \wedge [\pi_2]\phi$
$[\phi_1?]\psi \leftrightarrow (\phi_1 \rightarrow \psi)$
$[\pi^*]\phi \leftrightarrow \phi \wedge [\pi][\pi^*]\phi$

- The Induction Axiom $(\phi \wedge [\pi^*](\phi \rightarrow [\pi]\phi)) \rightarrow [\pi^*]\phi$.

PDL can derive all Hoare rules, while generalizing modal logic and relational algebra. All its theorems are valid, and there is a nice completeness proof. And in all this, *PDL* is *decidable*. *PDL* distils the "essence of computability": cf. Harel et al. 2000.

Process theories: the larger picture

The above suggests a study of *process equivalences*, as processes have many natural levels of detail. Also, the decidability of *PDL* suggests a balance between the *expressive power*

of a logic and the *computational complexity* of its laws. This balance is also crucial to understanding natural language. Finally, *PDL* only studies sequential operations on programs, while *parallel composition* is the reality in network computing. This deep subject is pursued in Lambda Calculus, a system close to Montague's work, and Process Algebra (Bergstra et al. 2001). Parallel computation is *simultaneous action*, which can also be modeled by games (Abramsky 2008). Interestingly, our linguistic performance, too, is a parallel composition of grasping meanings and engaging in discourse at the same time.

4 Dynamic Semantics of Natural Language Sentences

Many systems of interpretation highlight actions that deal with anaphora, temporality, and many other expressions. A major paradigm is Discourse Representation Theory (Kamp and Reyle 1993), but here we present an approach due to Groenendijk and Stokhof 1991.

Translation Lore

Dynamic predicate logic (*DPL*) "dynamifies" first-order logic. Here is why. One is usually taught some folklore to make first-order formulas fit actual linguistic forms:

1	A *man* came in. *He* whistled.	The two underlined phrases can co-refer.
2	* *No man* came in. *He* whistled.	The two underlined phrases cannot co-refer.
3	* *He* whistled. A *man* came in.	The two underlined phrases cannot co-refer.
4	If *a man* came in, *he* whistled.	The two underlined phrases can co-refer.

The obvious translation $\exists x\, Cx \wedge Wx$ for 1 does not give the right scope, and one uses a bracket trick: $\exists x\, (Cx \wedge Wx)$. The translation for 2: $\neg\exists x\, Cx \wedge Wx$ does give the right scope, the quantifier does not bind the free variable in Wx. The translation for 3: $Wx \wedge \neg\exists x\, Cx$, too, is correct. But the translation for 4: $\exists x\, Cx \rightarrow Wx$ has the wrong scope, and one uses brackets plus a quantifier-change (though the sentence has $\rightarrow$ as its main operator): $\forall x\, (Cx \rightarrow Wx)$.

Dynamifying Standard First-Order Semantics

DPL assigns dynamic meanings without tricks. It reinterprets first-order formulas ϕ as *evaluation procedures*, transition relations between assignments like with programs:

Atoms as tests
$\mathbf{M}, s_1, s_2 \models Px$ iff $s_1 = s_2$ *and* $I^{\mathbf{M}}(P)(s_1(x))$

Conjunction as composition
$\mathbf{M}, s_1, s_2 \models \phi \wedge \psi$ iff *there is* s_3 *with* $\mathbf{M}, s_1, s_3 \models \phi$ *and* $\mathbf{M}, s_3, s_2 \models \psi$

Negation as failure test
$\mathbf{M}, s_1, s_2 \models \neg\phi$ iff $s_1 = s_2$ *and there exists no* s_3 *with* $\mathbf{M}, s_1, s_3 \models \phi$

Existential quantification as random reset
$\mathbf{M}, s_1, s_2 \models \exists x$ iff $s_2 = s_1[x:=d]$ *for any object d in the domain.*

Example Dynamic evaluation and bindings explained.

(1) Evaluating $\exists x\, Cx \wedge Wx$ composes a random reset with two successive test actions. This moves from states s to states $s[x:=d]$ where both $C(d)$, $W(d)$ hold. (2) $Wx \wedge \exists x\, Cx$ composes a test, a random reset, and one more test. This moves from states s where $W(s(x))$ holds to states $s[x:=d]$ where $C(d)$ holds: no binding achieved. (3) The non-binding is explained by the negation test, which leaves no new value for x to co-refer. (4) To get the implications right, we define $\phi \rightarrow \psi$ as $\neg(\phi \wedge \neg\psi)$. This works out to a new test: *Every successful execution of ϕ can be followed by one for ψ.* This does what it should for both implications.

Logic as Evaluation Algebra

Conceptually, *DPL* makes predicate logic a theory of two basic actions: variable resets and atomic tests. "Standard logic" then becomes a mix of general relation algebra, at the level of the dynamic logic *PDL*, plus specific laws for reset actions on first-order models. Thus we see an intriguing fact. The *basic process logic* of evaluation is decidable—the undecidability of first-order logic arises from debatable special mathematical features of assignments (van Benthem 1996). *DPL* views also apply to discourse, suggesting notions of "dynamic inference": cf. the mentioned sources.

We leave the reader with a thought. If meaning is dynamic, what computational process drives natural language? Do typical program structures like *WHILE* and * iteration make sense?

5 Logical Dynamics of Conversation

We now move from sentences to discourse, and information flow in communication.

Example Cooperative questions and answers.

I ask you in Amsterdam: "Is this building the Rijksmuseum?" You answer: "Yes." This is a simple thing we all do all the time, but subtle information flows. By asking the question, I tell you that I do not know the answer, and that I think you may know. And by answering, you do not just convey a topographical fact—you also make me know that you know, and as a result, you know that I know that you know, etc.

Common knowledge, at every depth of iteration, mixes factual information and social information about what others know. The latter is the glue of communication, according to philosophers, economists, and psychologists. Hence we need logics that treat information flow with actors on a par. We do this by "dynamifying" the static logic of knowledge:

Epistemic Logic

The *epistemic language EL* extends propositional logic with modal operators $K_i\phi$ (*i knows that ϕ*), for each agent i in a total group I, and $C_G\phi$: *ϕ is common knowledge* in the subgroup G. The inductive syntax rule is as follows:

$$p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid C_G\phi$$

This language describes the Question/Answer scenario with formulas like

(i) $\neg K_{\mathbf{Q}}\varphi \wedge \neg K_{\mathbf{Q}}\neg\varphi$ (**Q** does not know whether φ),
(ii) $\neg K_{\mathbf{Q}}\neg(K_{\mathbf{A}}\varphi \vee K_{\mathbf{A}}\neg\varphi)$ (**Q** thinks that **A** may know the answer).

After communication, we have $K_{\mathbf{A}}\varphi \wedge K_{\mathbf{Q}}\varphi$, $K_{\mathbf{Q}}K_{\mathbf{A}}\varphi \wedge K_{\mathbf{A}}K_{\mathbf{Q}}\varphi$, and even $C_{(\mathbf{Q},\mathbf{A})}\varphi$.

Formally, consider models $\mathbf{M} = (W, \{\sim_i \mid i \in G\}, V)$, with worlds W, accessibility relations $\sim_i$ for agents $i \in G$ between worlds, and V a valuation as usual. *Pointed models* $(\mathbf{M}, s)$ have an actual world s for the true state of affairs (perhaps unknown to the agents). Here accessibility no longer encodes actions but *information ranges*: the options agents see for the actual world. Further conditions on $\sim_i$ encode special assumptions about agents' powers of observation and introspection. Very common are *equivalence relations*: reflexive, symmetric, and transitive. Such "information diagrams" interpret the epistemic language. Here are the key clauses:

$\mathbf{M}, s \models K_i\phi$	iff	*for all* t *with* $s \sim_i t$: $\mathbf{M}, t \models \phi$
$\mathbf{M}, s \models C_G\phi$	iff	*for all* t *that are reachable from* s *by some finite sequence of arbitrary* $\sim_i$ *steps* $(i \in G)$: $\mathbf{M}, t \models \phi$

We draw one model for a simple question answer episode (omitting reflexive arrows). Agent **Q** does not know if p, but **A** is informed about it:

In the actual world (the black dot), the following formulas are true: p, $K_{\mathbf{A}}p$, $\neg K_{\mathbf{Q}}p \wedge \neg K_{\mathbf{Q}}\neg p$, $K_{\mathbf{Q}}(K_{\mathbf{A}}p \vee K_{\mathbf{A}}\neg p)$, $C_{(\mathbf{Q},\mathbf{A})}(\neg K_{\mathbf{Q}}p \wedge \neg K_{\mathbf{Q}}\neg p)$, $C_{(\mathbf{Q},\mathbf{A})}(K_{\mathbf{A}}p \vee K_{\mathbf{A}}\neg p)$. This is a good reason for **Q** to ask **A** about p: He knows that she knows the answer.

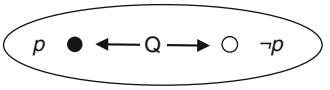


Figure 5.7.1

After the answer "Yes", intuitively, this changes to the following one-point model:

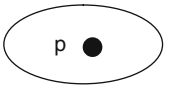


Figure 5.7.2

Now, common knowledge $C_{(\mathbf{Q},\mathbf{A})}p$ holds at the actual world. Epistemic logic sharpens intuitive distinctions about information, especially levels of group knowledge. Communication often turns *implicit* group knowledge into explicit knowledge.

Axiom Systems for Epistemic Inference

Complete logics capturing epistemic reasoning about oneself and others are known (Fagin et al. 1995). The base system is a minimal modal logic. Structural restrictions to equivalence relations add *S5* axioms of introspection, while the complete logic of common knowledge can be axiomatized with *PDL*-techniques.

A Dynamic Turn: Public Update by Elimination

Now for the logical dynamics of information flow. A pilot system for exploring this starts from a folklore view: An event *!P* yielding the information that *P* is true *shrinks the current model* to just those worlds that satisfy *P*. This is called *public hard information*. More precisely, for any epistemic model **M**, world *s*, and *P* true at *s*, the model *(***M** | *P, s)* (‘**M** *relativized to P at s*’) is the submodel of **M** whose domain is the set of worlds *{t∈***M** | **M**, *t* ⊨*P}*. Drawn in a simple picture, an update step then goes

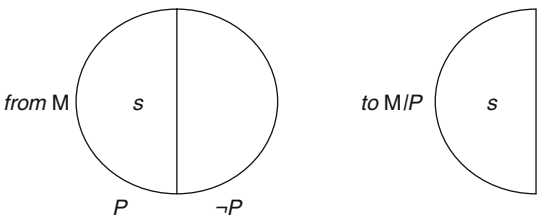


Figure 5.7.3

This mechanism models communication, but also acts of public *observation*. It has been applied to games and other social scenarios. Crucially, truth values of formulas may change after update: agents who did not know that *P* now do. We need a logic to keep things straight.

Dynamic Logic of Public Announcement

The language of *public announcement logic PAL* adds action expressions to *EL*, plus matching dynamic modalities, defined by the syntax rules:

Formulas P: $p \mid \neg\phi \mid \phi\vee\psi \mid K_i\phi \mid C_G\phi \mid [A]\phi$
Action expressions A: $!P$

Here the semantic clause for the dynamic action modality “looks ahead” between different models:

M, *s* ⊨ *[!P]*ϕ iff *if* **M**, *s* ⊨ *P*, *then* **M** | *P*, *s* ⊨ ϕ

A typical assertion here is *[!P]*$K_i\phi$*,* which states what agent *i* will know after receiving hard information that *P* is the case. Reasoning about information flow revolves around a dynamic *recursion equation* that relates new knowledge to old knowledge an agent had before:

The following equivalence is valid for *PAL*: $[!P]K_i\phi \leftrightarrow (P \rightarrow K_i(P \rightarrow [!P]\phi))$.

The reader may find it helpful to prove this. The complete and decidable logic for knowledge under public communication is well understood. *PAL* is axiomatized by any complete epistemic logic over static models plus *recursion axioms*

$$
\begin{array}{lcll}
[!P]q & \leftrightarrow & P \rightarrow q & \textit{for atomic facts } q \\
[!P]\neg\phi & \leftrightarrow & P \rightarrow \neg[!P]\phi & \\
[!P](\phi \wedge \psi) & \leftrightarrow & [!P]\phi \wedge [!P]\psi & \\
[!P]K_i\phi & \leftrightarrow & P \rightarrow K_i(P \rightarrow [!P]\phi) &
\end{array}
$$

There is more here than meets the eye. The logic *PAL* uncovers many subtleties of natural language. Suppose that in our question-answer episode, **A** had not said $!P$, but the equally true "You don't know that P, but it is true" ($\neg K_Q P \wedge P$). The latter "Moore sentence" achieves the same update, but it has become *false* afterward! Statements switching their own truth values are essential in understanding conversation, puzzles, and games (Geanakoplos and Polemarchakis 1982). For much more about *PAL* and related systems, including links to epistemology, cf. Baltag et al. 1998, van Ditmarsch et al. 2007, van Benthem 2011.

Compared to dynamic semantics, logical dynamics of information flow has a discourse focus. This is "pragmatics"—but from a logical point of view, the border with semantics is thin.

6 The Logical Dynamics of Agency

Public announcement is just the start of a dynamics of interactive agency. We mention a few dimensions of a richer picture of what language users are and do.

From Knowledge to Belief

Language users do not just have knowledge, but also *beliefs*. What they hear involves belief revision (Gaerdenfors 1988), and this process can be triggered by information that is "soft" rather than hard, depending on the reliability of the source. Dynamic logics for belief revision "dynamify" static doxastic logics, where agents believe what is true in their "most plausible" worlds. This time, update does not eliminate worlds: It transforms the *relative plausibility* that agents assign to worlds. Forming and correcting beliefs is a learning ability that is more essential to human intelligence than just recording hard information. Rationality is not being correct all the time, but having a talent for *correction*.

Private Information

Information flow is driven by a differential: we do not all know the same things. Dynamic epistemic logics can also model *private* communication in a group (think of emails with *bcc* rather than *cc*), a phenomenon of high complexity. They even deal with lying and cheating, a central topic in real language, since most communication is unreliable to some extent. The usual Gricean focus on helpful truthful communication seems otherwordly.

Questions and Issue Management

Questions do not just convey information, they also *direct discourse* by raising and modifying topics. This is crucial to language, communication and inquiry. A logical

dynamics of questions must represent *issues* and actions modifying these. Two recent flavors are Groenendijk 2009, van Benthem and Minica 2009.

Preference and Evaluation Dynamics

Agency is not just information dynamics. Rational decision and strategic interaction involve a balance of information and *evaluation*, encoded in our preferences. Entanglement of information and evaluation, and preference change pervade deontic logic (Gruene Yanoff and Hansen 2010) and games. But again natural language remains close. We often change our evaluation of situations by speech acts such as suggestions or *commands* from some moral or esthetic authority. Philosophers have also drawn attention to the normative character of "discourse obligations" (Brandom 1994).

And Beyond

Further relevant features of agency studied in logic include *trust*, *intentions* and *commitments*, but we the picture should be clear by now. Dynamic analysis of language may at the same time have to be an analysis of the agents using that language.

7 Games in Logic and Natural Language

Beyond logics of agency lies a world of *games*. Dynamic logics describe single update steps, but the next level of language use is strategic behavior over time. What I say now responds to what you say, but it is usually directed toward a future goal, and part of a long-term plan. We also saw that language involves iterated social knowledge of agents about each other and multi-agent equilibrium. Both have received sophisticated treatments in game theory.

Evaluation Games and "Game-Theoretic Semantics"

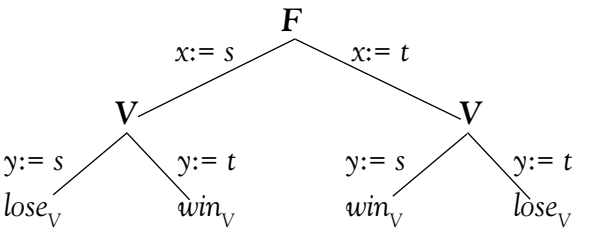
We illustrate games for natural language with a simple pilot system, again in terms of first-order logic (Hintikka and Sandu 1997). Evaluation of formulas ϕ can be cast as a game for two players. *Verifier* **V** claims that ϕ is true in a model **M**, *s*, *Falsifier* **F** that it is false. Here are natural moves of defense and attack:

atoms *Pd*, *Rde*, . . .	**V** wins if the atom is true, **F** if it is false
disjunction $\phi \vee \psi$	**V** *chooses* which disjunct to play
conjunction $\phi \wedge \psi$	**F** *chooses* which conjunct to play
negation $\neg\phi$	*role switch* between the players, play continues with ϕ

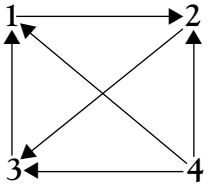
Next, the quantifiers make players look inside **M**'s domain of objects:

$\exists x\, \phi(x)$	**V** *picks an object d*, and play continues with $\phi(d)$
$\forall x\, \phi(x)$	the same move, but now for **F**

The game schedule is determined by the form of ϕ. Consider a model **M** with two objects *s*, *t*. Here is the complete game for the formula $\forall x\, \exists y\, x \neq y$ in **M**, pictured as a tree of possible moves:



This is a game of perfect information: players know throughout what has happened. Branches are the possible plays, with 2 wins for each player. But **V** is the player with a *winning strategy*, she has a rule for always winning. For a more exciting example, look at the following network with arrows for directed communication links (all self-loops are present but not drawn):



The formula $\forall x\ \forall y\ (Rxy \vee \exists z(Rxz \wedge Rzy))$ says that any two nodes can communicate in two steps. Just analyze the game and see who can win.

Logic and Games

The crucial fact about evaluation games is this equivalence:

A formula is true iff Verifier has a winning strategy,

while it is false iff Falsifier has a winning strategy. This follows from *Zermelo's Theorem* on "determined games,", a stepping stone toward solution procedures for games with richer preferences than winning and losing ("Backward Induction"). In evaluation games, logical constants change from "control expressions" for procedures to *game actions* like making a choice or performing a role switch. Thus, at the heart of natural language, there is a multi-agent game algebra of its users. Actually, in game theory, the norm is *imperfect information*: Players need not know exactly where they are in a game tree (think of card games). Hintikka and Sandu 1997 claim that "branching patterns" of independence between quantifiers in natural langage involve imperfect information about objects chosen by one's opponent.

Evaluation games are not a realistic account of discourse, which is usually about consistency rather than truth. But modern logic has many further games that are relevant.

"Logic Games"

Games in logic analyze argumentation (Lorenzen 1955), compare models (Ehrenfeucht 1957), construct models (Hodges 1985), etc. In each case, winning strategies for various agents encode basic notions. A winning argumentation strategy is a proof for the claim,

or if the opponent wins, a counter-model. Games are at the heart of modern logic, and quantification is deeply tied to dependent action.

Signaling Games for Meaning

A quite different use of games in language has emerged in Parikh 2000, Jaeger and van Rooij 2007, and Gaerdenfors and Warglien 2006. These start from the *signaling games* in Lewis 1969 that analyze basic lexical meanings. We have situations and linguistic objects that can represent them. Agents might choose any association, but stable conventions are Nash Equilibria in a game where a Sender chooses a coding scheme and Receiver a decoding, with some plausible assumptions on their utility functions. Thus meanings become *equilibria in language games*. Richer infinite "evolutionary games" can even explain diachronic phenomena, or emergence of linguistic conventions, using thought experiments in terms of fitness and stability against invaders.

Integrating Different Games

Signaling games are very different from logic games, where meanings are given. Integration of these perspectives on natural language is an open problem.

Logic and Game Theory

We have discussed special games for linguistic and logical activities. But there is also an interface of logic and general game theory, in the study of strategies, information, and reasoning of agents. This involves epistemic, doxastic, and dynamic logics for analyzing rational play and game solution. This relates to the sense in which computer scientists have embraced games in *multi-agent systems* (Shoham and Leighton-Brown 2008), and philosophers in epistemology (Stalnaker 1999). While this interface is not disjoint from language and logic games, we will not pursue it here: cf. van der Hoek and Pauly 2006 for this area of investigation.

Coda: Temporal Perspective

Lexical meaning assignment, evaluation, or argumentation are special-purpose short-term processes. These run against the backdrop of an infinite linguistic process over time: the "operating system" of natural language. Here language dynamics meets with temporal logics (Parikh and Ramanujam 2003, Belnap et al. 2001), learning theory (Kelly 1996), and infinite computational processes (Grädel et al. 2003).

8 Discussion: Putting the Dynamics Together

Dynamic Semantics Versus Dynamic Logic

Dynamics in this chapter has two different strands. "Dynamic semantics" is a *new account of meaning*, replacing truth-conditional accounts, and generating "nonstandard logics." By contrast, dynamic logics of information update keep the old language with its semantics and logic, but *add dynamic events* as a new layer. The former approach is implicit: The dynamics "loads" the meaning of the old language, while the latter approach is

explicit, the dynamics occurs in operators extending the old language. The implicit/explicit contrast occurs widely in logic—but what fits natural language best?

Combined Architectures

How can we turn our carousel of dynamic activities and games into one integrated story of language use? Combining logics can be tricky. Simple decidable logics for knowledge and time combine into highly undecidable logics for agents with perfect memory (Halpern and Vardi 1989). No integration is known for dynamics, and we may first need an account of "linguistic agents", the way Turing analyzed "computing agents".

Cognitive Realities

Finally, natural language is an interface where logic meets reality—and so dynamic logics meet cognitive science. Van Benthem 2010 proposes studying language in a broad sense here, including "successful insertions" of new logic-inspired behavior.

9 Conclusion

We have shown how natural language meets with dynamic logics of meaning and agency, leading to new interfaces between logic, philosophy, linguistics, computer science and game theory.

References

Abramsky, S. (2008) 'Information, Processes and Games', in P. Adriaans and J. van Benthem (eds) *Handbook of the Philosophy of Information*, Amsterdam: Elsevier, 483–549.

Baltag, A, L. Moss and S. Solecki (1998) "The Logic of Public Announcements, Common Knowledge and Private Suspicions," in *Proceedings TARK 1998*, Los Altos: Morgan Kaufmann, 43–56.

Belnap, N., M. Perloff, and M. Xu (2001) *Facing the Future*, Oxford: Oxford University Press.

Bergstra, J., A. Ponse and S. Smolka (eds.) (2001) *Handbook of Process Algebra*, Amsterdam: Elsevier.

Blackburn, P., M. de Rijke, and Y. Venema (2000) *Modal Logic*, Cambridge: Cambridge University Press.

Brandom, R. (1994) *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, MA: Harvard University Press.

Dekker, P. (2008) "A Guide to Dynamic Semantics," Amsterdam: Institute for Logic, Language and Computation.

Ehrenfeucht, A. (1957) "Application of Games to Some Problems of Mathematical Logic," *Bulletin Polish Academy of Sciences* Cl. III, 35–37.

Fagin, R., J. Halpern, M. Vardi, and Y. Moses (1995) *Reasoning about Knowledge*, Cambridge, MA: MIT Press.

Gaerdenfors, P. (1988) *Knowledge in Flux*, Cambridge, MA: MIT Press.

Gaerdenfors, P. and M. Warglien (2006) "Cooperation, Conceptual Spaces, and the Evolution of Semantics," in: P. Vogt et al. (eds.) *Symbol Grounding and Beyond*, Heidelberg: Springer, 16–30.

Geanakoplos, J. and H. Polemarchakis (1982) "We Can't Disagree Forever," *Journal of Economic Theory* 28, 192–200.

Grädel, E., W. Thomas and Th. Wilke (eds.) (2003) *Automata, Logics, and Infinite Games*, Heidelberg: Springer.

Groenendijk, J. (2009) "Inquisitive Semantics: Two Possibilities for Disjunction," in P. Bosch, D. Gabelaia and J. Lang (eds.) *Proceedings Seventh Tbilisi Symposium on Language, Logic and Computation*, Heidelberg: Springer, 80–94.

Groenendijk, J. and M. Stokhof (1991) "Dynamic Predicate Logic," *Linguistics and Philosophy* 14, 39–100.

Groenendijk, J., M. Stokhof, and F. Veltman (1996) "Coreference and Modality," in S. Lappin (ed.) *Handbook of Contemporary Semantic Theory*, Oxford: Blackwell, 179–216.
Grüne Yanoff, T. and S.-O. Hansen (2010) *Preference Change*, Dordrecht: Springer.
Halpern, J. and M. Vardi (1989) "The Complexity of Reasoning about Knowledge and Time," *Journal of Computer and System Sciences* 38, 195–237.
Harel, D. (1987) *Algorithmics: The Spirit of Computing*, Reading, MA: Addison-Wesley.
Harel, D., D. Kozen, and J. Tiuryn (2000) *Dynamic Logic*, Cambridge, MA: MIT Press.
Hintikka, J. and G. Sandu (1997) "Game-Theoretical Semantics,",in: J. van Benthem and A. ter Meulen (eds.) *Handbook of Logic and Language*, 361–410.
Hodges, W. (1985) *Building Models by Games*, Cambridge: Cambridge University Press.
Jaeger, G and R. van Rooij (2007) "Language Structure: Psychological and Social Constraints," *Synthese* 159, 99–130.
Kamp, H. and U. Reyle (1993) *From Discourse to Logic,* Dordrecht: Kluwer.
Kamp, H. and M. Stokhof (2008) "Information in Natural Language," in P. Adriaans and J. van Benthem (eds.) *Handbook of the Philosophy of Information*, Amsterdam: Elsevier, 49–111.
Kelly, K. (1996) *The Logic of Reliable Inquiry*, Oxford: Oxford University Press.
Lewis, D. (1969) *Convention*, Cambridge, MA: Harvard University Press.
Lewis, D. (1979) "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8, 339–359.
Lorenzen, P. (1955) *Einführung in die Operative Logik und Mathematik*, Berlin: Springer Verlag.
Meulen, A. ter (1995) *Representing Time in Natural Language: The Dynamic Interpretation of Tense and Aspect*, Cambridge, MA: MIT Press.
Montague, R. (1974) *Formal Philosophy*, New Haven CT: Yale University Press (edited by R. H. Thomason).
Parikh, P. (2000) *The Use of Language*, Stanford: CSLI Publications.
Parikh R. and R. Ramanujam (2003) "A Knowledge-Based Semantics of Messages,"*Journal of Logic, Language and Information* 12, 453–467.
Searle, J. and D. Vanderveken (1985) *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press.
Shoham Y. and K. Leyton-Brown (2008) *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations,* Cambridge: Cambridge University Press.
Stalnaker, R. (1978) "Assertion," in P. Cole (ed.) *Syntax and Semantics* 9, New York: Academic Press, 315–32.
Stalnaker, R. (1999) "Extensive and Strategic Form: Games and Models for Games,"*Research in Economics* 53, 293–291.
Van Benthem, J. (1996) *Exploring Logical Dynamics*, Stanford: CSLI Publications.
Van Benthem, J. (1999) *Logic in Games*, Amsterdam: Institute for Logic, Language and Computation. To appear in *Texts in Logic and Games*, Springer.
Van Benthem, J. (2011) *Logical Dynamics of Information and Interaction*, Cambridge: Cambridge University Press.
Van Benthem, J. and S. Minica (2009) "Dynamic Logic of Questions," in J. Horty and E. Pacuit (eds.), *Proceedings LORI-2 Chongqing*, Heidelberg: Springer Publishers, 28-42. To appear in *Journal of Philosophical Logic*, 2011.
Van Benthem, J. and A. ter Meulen, eds. (1997) *Handbook of Logic and Language*, Amsterdam: Elsevier Science Publishers.
van Ditmarsch, H., W. van der Hoek and B. Kooi (2007) *Dynamic Epistemic Logic*, Dordrecht: Springer.
Van der Hoek, W. and M. Pauly (2006) "Modal Logic for Games and Information,", in P. Blackburn, J. van Benthem, and F. Wolter (eds.) *Handbook of Modal Logic*, Amsterdam: Elsevier Science Publishers, 1077–1148.
Van Leeuwen, J. (ed.) (1990) *Handbook of Theoretical Computer Science*, Amsterdam: Elsevier.
Veltman, F. (1996), "Defaults in Update Semantics", *Journal of Philosophical Logic* 25, 221–261.

5.8
INTUITIONISM

Allen Hazen

1

Intuitionism is a viewpoint in the philosophy of mathematics, developed by the Dutch mathematician L.E.J. Brouwer, which rejects many standard ("classical") mathematical results and methods, and in particular rejects as invalid some principles of standard ("classical," two-valued) elementary logic. It is held by a minority of mathematicians (a minority even of those willing to go public with philosophical statements) but has inspired much mathematical research: the term 'intuitionism' is also used for the body of mathematics developed in accordance with the strictures of the philosophy, a body by now of sufficient mathematical richness to be of interest even to researchers not concerned with philosophical issues. The term 'constructivism' is also used for similar philosophical viewpoints and mathematical methods: when a contrast is made between them, 'intuitionism' refers to a viewpoint or a set of mathematical methods more like those of Brouwer himself and 'constructivism' more generically to these and related ones. 'Intuitionistic (or 'constructive') logic' refers to a specific version of propositional and first-order logic first formalized by Brouwer's younger colleague Arend Heyting. There doesn't seem to be any useful analogy with the views known as intuitionistic in such areas of philosophy as meta-ethics, nor is constructive mathematics connected to what is called 'social constructivism' about mathematics and mathematics education.

The mathematical notion that is perhaps most central to the motivation of the intuitionistic philosophy and constructivist mathematics is that of a constructive proof, which can be illustrated with elementary geometrical examples. Typical existence proofs in Euclidean geometry are given by explicit construction. To show that, for example, for any triangle there is a circle passing through its three vertices, one specifies a sequence of operations with (idealized) straight-edge and compass which will result in the drawing of exactly this circle. Generalizing: to prove, constructively, that a certain mathematical object exists, you present it, or at least present a method (the construction) which will, provably, find it. For an example of a nonconstructive proof, consider the theorem that the closed unit interval is sequence compact: for any infinite sequence of points in [0,1], there exists a subsequence (i.e. an infinite sequence of points all occurring, in the same order though perhaps with gaps, in the original sequence) that converges to a point in the interval. Proof: divide the interval into two (overlapping) subintervals, [0,.5] and [.5,1], each of length a half: at least one will contain infinitely many points of the sequence. Now divide this subinterval (or, say, the left one if both contain infinitely many points of the sequence), and divide it into two subintervals of length a quarter. . . . Thus there is an infinite sequence of successively shorter nested closed intervals

(the n-th being of length $1/2^n$): their intersection contains a single point. Now consider the subsequence of the original sequence of points whose first point is the first from the original sequence lying in the chosen half, whose second is the next point in the original sequence lying in the chosen quarter, and so on: this subsequence converges to the point in the intersection of the intervals. This is a perfectly acceptable proof in standard mathematics, but it gives no hint as to where the point of convergence is! Even if the original sequence of points is specified in some explicit manner (say by presenting an algorithm computing their location one after another), there may be no equally explicit way of displaying the converging subsequence or locating the point of convergence. (If giving an algorithm is our standard for explicit specification, this follows immediately from the unsolvability of the Halting Problem for Turing Machines.)

There is an obvious sense in which a constructive proof is more informative than a non-constructive. Quite independent of any philosophical commitments, a mathematician might *prefer* constructive proofs, might find nonconstructive proofs *unsatisfying*, might even feel that a nonconstructive proof was somehow "not cricket" as mathematics: thus Paul Gordan's legendary "This is not mathematics! This is theology!" in response to a nonconstructive existence proof by David Hilbert. (The context and intention of Gordan's supposed exclamation aren't clear; the saying has become legendary, though, because it is felt to express an attitude that was more than idiosyncratic.)

Brouwer was not interested in formal logic but saw that if nonconstructive proofs are rejected as invalid, then some traditional principles of logic are "untrustworthy." Proof of the negation of an assumption by *reductio* (showing that from the assumption absurdity can be derived) was all right: how else are you to prove a negation? The closely related *indirect proof*, however—inferring an arbitrary, not necessarily negative, conclusion when absurdity is derived from the assumption of its negation—was not: using indirect proof to establish an existence theorem is the archetype of the nonconstructive. Other principles could be rejected because if accepted they would license indirect proof. Thus *double negation elimination* (from Not-Not-A to infer A) would give indirect proof with one extra step: use *reductio* to establish the negation of the negation of an existence claim, and then *d.n.e.* Similarly for the Law of Excluded Middle (either A or Not-A, for an arbitrary proposition A): it would allow us to argue "Either the existence claim is true or it isn't; the latter is absurd, so the former must hold." (Note how the *L.E.M.* is assumed in the nonconstructive proof of sequence compactness above: to show that, an arbitrary interval containing infinitely many points of the sequence, either its left half or its right must contain infinitely many, we must assume of each half that either it contains infinitely many or it doesn't contain infinitely many.) Addition of any of these principles to intuitionistic logic yields a formulation of classical, two-valued, logic.

A subtler, but perhaps more fundamental, point involves quantifiers. Suppose you have established that for every object x in the domain, either A holds or Fx, where A is some statement not containing the variable x. (Typical examples of this form arise when F is a "decidable" predicate, one for which it holds for every x that either Fx or Not-Fx: let A be the statement that there exists some y for which Not-Fy.) Classically you can infer that either A holds or it holds that for every x Fx. (This is the converse of one of the prenexing operations, moving a universal quantifier into one disjunct of a disjunction.) Intuitionistically you can't: statements with embedded quantifiers are not in general equivalent, intuitionistically, to prenex statements. As with the three principles mentioned above, if this inference were allowed it would license some

nonconstructive proofs: where F is a decidable predicate, it would allow a nonconstructive proof of the existence of a non-F object. It can, however, be added to intuitionistic first-order logic without affecting the purely propositional fragment of the logic; the resulting logic, intermediate in strength between intuitionistic and classical, is often called "constant domain" (for short: CD) logic, as it is characterized in Kripke's semantics by models with a constant domain across "possible worlds".

There is actually some debate about the status of this last principle among constructivist mathematicians. In my rough and informal initial characterization of constructive proof I said that presenting an algorithm that will (provably) find an example counts as a constructive proof of an existence theorem. Now let F be a decidable predicate of natural numbers, in the sense that there is an algorithm for deciding, with respect to a given natural number, whether F holds of it or not. (Intuitionists would take any proof that such an algorithm works as establishing that, for every natural number n, either Fn or Not-Fn, so any predicate decidable in this sense is also decidable in the sense suggested in the previous paragraph.) Then if it is proven (by reductio) that it is not the case that for every n, Fn, we could (and a classical recursive function theorist would) say that there *is* an algorithm for finding a non-F number: just test the natural numbers, one after another, for F-ness until the first counterexample appears! Some mathematical constructivists (in particular the Russian school stemming from the work of A. Markov) would count this as a constructivistically acceptable proof of the existence of a non-F number; Brouwerian intuitionists wouldn't. Presenting an algorithm doesn't, for the stricter constructivists like Brouwer, count as a constructive proof unless the proof the algorithm will work is itself constructive.

2

One could feel that constructive proofs are more interesting than nonconstructive, or deem them more valuable, or even argue that mathematics journals should not waste their page space by publishing nonconstructive proofs, without holding that they are literally invalid, that they fail to establish the truth of their conclusions. It is this claim that makes intuitionism or constructivism interesting for philosophers not specifically concerned with the methodology of mathematics, but it is not obvious how it should be justified, and Brouwer's own exposition is on the Delphic side. We can approach the issue by considering disjunctive statements, such as instances of the law of excluded middle. As a starting point, on the natural interpretation, a disjunction is true just in case at least one of its disjuncts is. (As the many proposals for the use of Supervaluations in semantics show, this is a defeasible principle, but I will take it as a starting point in this chapter.) Similarly, an existential quantification is naturally thought of as true only if it has some true instance (though here there is the complication that the "witness" verifying the existential quantification might not have a name). If we can prove that one of the disjuncts is true, then we may obviously infer that the disjunction is; this principle is shared by intuitionistic and classical logic. In the other direction, classical logic holds some disjunctions to be true, or even logically valid, even though neither disjunct is provable: the law of excluded middle holds even for propositions which are neither provable nor refutable. Intuitionistic logic is more circumspect. No disjunction is a logical theorem of intuitionistic logic unless at least one of its disjuncts is, and, similarly, no existential quantification is unless some specific instance is provable. Nor can a disjunctive (or existential) conclusion be derived unless one or more premisses are of

disjunctive or existential character (in a sense that is a bit picky to make precise but intuitively clear). Nor is this a matter of logic alone: natural formal axiomatizations of intuitionistic mathematical theories (including, centrally, what is called "Heyting Arithmetic": essentially first-order Peano Arithmetic with intuitionistic logic) permit the proof of disjunctive (existential) theorems only when disjuncts (instances) are provable.

These results about intuitionistic formal systems are striking, and are often cited as confirming the "constructive" character of the systems, but they suggest something more. Particular formal axiomatizations cannot be taken as exhaustive of the notion of intutitionistic provablility, any more than a particular axiomatization exhausts that of provability in classical mathematics. (Gödel's incompleteness theorems carry over directly. The consistency of Heyting Arithmetic is unprovable in Heyting Arithmetic, just as the consistency of classical P.A. is unprovable in P.A., but would be affirmed by intuitionists just as classical mathematicians would affirm the consistency of P.A.) Still, it is a tempting conjecture that "real" intuitionistic provability satisfies the same "constructivity" criteria: disjunctions provable *only* when disjuncts are, existential quantifications only when instances are. What this suggests is that truth and proof are more intimately connected than they are on the classical picture. For a classical logician, proof is a valuable but not always applicable way of coming to know that a proposition is true, but the truth of the proposition doesn't depend on the existence of a proof. (Every instance of the law of excluded middle has at least one true disjunct, but in many cases no proof of either disjunct is available.) The intuitionist, however, suggests that no proposition is true unless it has a proof: perhaps, in some sense, it is the proof that *makes* it true.

One way of interpreting this close relationship between truth and proof is, in a broad sense, *epistemological*. A number of philosophical commentators on intuitionism, most notably Michael Dummett, have argued that what we understand in understanding a mathematical proposition involves essentially a grasp of how, in princple, it could be proven. As a result it does not make sense to attribute truth to a mathematical proposition unless we could, in principle, come to know it: could prove it. Some, Dummmett again being the most prominent, have attempted to generalize this conception beyond mathematics, claiming that the understanding of any proposition depends on a grasp of how it can be verified, and that truth can, therefore, never "transcend" what we could in principle come to know. This has led to what is known as "Anti-Realistic" or "Assertibility" semantics: probably this is the main reason why philosophers of language have taken as much interest in intuitionism in recent decades as they have. Inevitably, given the attempted generalization, much of the literature in the philosophy of language is not specifically concerned with mathematical intuitionism or the ways in which mathematical theorizing and subject matter are special.

Rather than attempting to summarize this literature, I would like to describe a few features of intutitionistic mathematics and logic that seem to me of particular philosophical interest. The broadly epistemological interpretation just alluded to, focusing on what it is to understand a mathematical proposition, is attractive in many ways, particularly in its recognition of the centrality of the notion of (constructive) *proof* in intuitionism. A different, more ontological, interpretation is, however, also possible, and I think gives a simpler way of understanding the peculiarities of intuitionistic mathematics and of motivating intuitionistic logic.

Mathematical theorizing, at least in its superficial form, appears to be about abstract entities having, in general, no physical (or psychological) embodiment. To take the

simplest example, the natural number series is infinite: after every natural number there is another. Many, many, philosophers have been sceptical of the existence of such entities. Whitehead and Russell, for example admitted that their Axiom of Infinity (needed in *Principia Mathematica* to guarantee the endlessness of the number series) was dubious, and Dedekind's ingenious attempt to prove the existence of an infinite "system" by citing a series of ideas tends to be dismissed nowadays, often with sarcastic comments about how even Dedekind couldn't hold infinitely many thoughts in mind simultaneously! We have, however, clear enough ideas about how, for example, the number series should go on that we can have a principled mathematical theory about it even if it doesn't exist. One might, and on the ontological interpretation intuitionists do, hold that reasoning about "imaginary" entities should be governed by a different logic from reasoning about things we are confident really exist. In reasoning about real things, the law of excluded middle is unobjectionable: one disjunct or the other of an instance of the law will be true in virtue of the facts about the objects referred to, even if we don't (or can't) know which. When the objects we reason about don't exist, this justification of the law of excluded middle is no longer available. It is plausible that the "truth" about nonexistent objects is simply whatever follows from our conception of them, so that it is appropriate to use a logic which allows truth and provability to coincide. This point of view, I would argue, motivates intuitionistic logic as well as the epistemological interpretation, and, indeed, motivates it for "liberalized" forms of constructivism which are not covered by the epistemological interpretation.

3

For Brouwer *some* of the objects of mathematics, namely the "mental constructions" someone has already consciously effected, are real. If one clearly perceives (with the mind's eye) some finite pattern—the solution to a chess problem, for example—then it is a real thing about which we can reason classically. Over time we can effect more constructions: create new patterns. (A platonist, holding that forms have an abstract existence independent of any physical or mental realization, would say we don't really *create* patterns, but intuitionists are not platonists.) The whole infinite set of all possible finite patterns, however, does not exist. The universe the mathematician reasons about, then, can be thought of as, not completed, but *under construction*. So, how should we reason about something that is still under construction?

A nonmathematical example can illustrate the principle. Suppose you are an ethical property developer: you want to sell properties in a new suburb your company is developing, but you also want your sales representatives to be scrupulously honest when they describe them to potential buyers. Honesty, however, doesn't involve describing the current state of the development: potential buyers aren't interested in the bulldozers currently at work, but rather in what the neighborhood will look like when they move in. What the sales representatives must do, then, is to describe things (honestly) in language appropriate to the description of an already-built suburb, even though much of it is still unbuilt, and though some details will depend on decisions not yet made. Obviously, anything that has already been built can be described: the street pattern, perhaps, and the model houses that have been built on one block, but we will want to go beyond that. The fundamental constraint is that the sales representative should not say anything that might be falsified later in the construction process.

On the basis of these considerations we can formulate a recursive definition of what will count as a true ("honest") description of an unbuilt neighborhood. As a base case we take the principle that what has already been completed can be described. The recursive clauses for conjunction, disjunction, and existential quantification are completely standard: a statement of one of these forms is true just in case both of its conjuncts or at least one of its disjuncts is true or there is an object in the domain verifying its matrix. The remaining logical operators cannot have their standard interpretations if the truth of a statement is to be robust against future development. Thus, for example, the sales representatives can't assert a negation, such as "No building on this block is over three stories," simply because no tall buildings have been placed there yet; there must be some guarantee that none *will* be built (for example. a prohibition in the building code, or perhaps the fact that two-story houses have already been built on all the sites on the relevant block). Similarly for conditionals; not only must it not be the case that construction so far verifies the antecedent but not the consequent, but there must be a guarantee that if, at any later stage of the development, the antecedent comes to be true, the consequent will come true at the same time (or have already come true when the antecedent does). As for universal quantifications (for example "Every building will be compliant, not only with the current fire code but with the strengthening being considered by the county"), not only must the objects already built in the domain satisfy the matrix, but there must be a guarantee that any new ones added will too.

It's hard to predict all the questions potential home-buyers may ask: we, as developers, can't just issue our sales representatives to read from the documentation we provide in response. They will have to reason, to deduce answers. And the standard of validity for their reasoning ought to be truth preservation, in the sense of truth just described. The honest deductive logic for them to use is intuitionistic logic. (As should be obvious to any expert reading this, the fantasy about ethical real estate development is simply a way of dressing up the standard Kripke-model semantics for intutionistic logic.)

The ontological interpretation of mathematical constructivism, then, takes it to be a sort of *semi-fictionalist* account of mathematics. Some mathematical objects can be treated as real existents, but in theorizing about them we put them into the context of a larger system of objects, most of which exist "only in the mind." The full system of mathematical objects treated in a given branch of mathematics is "under construction." But, whereas a housing development may actually be completed at some stage, a mathematical universe will *always* be under construction.

The description so far has been abstract: the example was non-mathematical, and not much has been said about how the distinction between real and fictitious entities applies in actual constructivist mathematics. It turns out that there are several options available here, and different mathematicians and philosophers of mathematics have not all found the same option attractive. For Brouwer, and probably for the majority of the minority of mathematicians who consider themselves constructivists, finite objects—natural numbers, or the sequences of moves in chess games—can be considered real. In his first publication rejecting classical logic ("On the untrustworthiness of logical principles," 1908) he noted that two-valued logic *was* appropriate in reasoning about a finite domain (at least if the atomic predicates used are decidable). This can be shown formally in intuitionistic arithmetic. The condition on decidable predicates is met—an atomic sentence of arithmetic is just a numerical equation whose truth value can be calculated by grade-school algorithms—and it is provable in Heyting

Arithmetic that the instances of the law of excluded middle for statements involving only bounded quantifiers (for all x less than t/there exists an x less than t) hold. The infinite series of the natural numbers, however, cannot be completed, and so is not real. Hence excluded middle ceases to hold when unbounded quantifiers (for all x/there exists an x) occur.

Others, however, are attracted by a less restrictive conception. Daniel Velleman (noting similarities between his position and those of such early twentieth-century "semi-intuitionists" as Baire, Borel, and Lebesgue) has argued that a denumerable structure, such as the natural number series, can be considered real and reasoned about classically, but that nondenumerable sets, such as that of real numbers, shouldn't be. (He also notes that a position like this, called "denumerabilism," was described, though not endorsed, by Robert McNaughton.)

Even further from Brouwer's position is the "liberal intuitionism" proposed by Lawrence Pozsgay as a philosophy of set theory. Here sets can be treated as real, and classical logic used when all quantifiers are restricted to range over the members of some set, but the full set-theoretic universe is treated as "under construction," and reasoned about in intuitionistic logic.

Real human beings, of course, are severely limited in what they can understand or remember or visualize: to use Descartes's example, we don't distinguish a thousand-sided figure from a ten-thousand-sided by forming distinct visual images. The connection of any sort of constructive mathematics with actual human cognitive capacity is thus tenuous: an epistemological motivation must appeal to knowability by idealized human minds. The stricter versions of constructivism require less idealization, but it is hard to see the kind of mind that could visualize arbitrary denumerable sets, or arbitrary sets, as even an idealization of our own.

4

Thinking of a formalized theory as describing, not a fully realized system of actually existent objects but something still under construction with possibilities of being extended in different ways, puts us in the domain of modal logic. The idea of interpreting intuitionistic logic in some sort of modal logic is much older than the model-theoretic semantics of intuitionistic logic; it seems plausible to conjecture that Kripke was led to his analysis of intuitionistic logic by putting his own slightly earlier model-theoretic treatment of modal logic together with the already familiar modal translations of intuitionistic logic.

Gödel announced the possibility of a translation of intutionistic propositional logic in a (classical) modal logic in an abstract of 1933. In it he stated that, under any of a variety of translation schemes, all (and, he correctly conjectured, only) the intuitionistically valid formulas of propositional logic translate into theorems of the modal logic S4: subsequent research has shown that (not always with the same translation schemes) many other modal logics can be substituted for S4. He suggested that the necessity operator could be thought of as expressing the notion *it is provable that*, noting that, in view of the Incompleteness Theorem, this had to be understood in terms of informal provability and not provability in a particular axiomatic system. (Modal logics whose necessity operators could be interpreted in terms of provability in a formalized system were discovered late and Goldblatt showed that such a "provability logic" could be used to interpret intuitionistic logic.)

The connection with Kripke's semantics for intuitionistic logic is clearest with a translation scheme slightly different form the ones Gödel considered. We can define it (extending it to full first-order logic) recursively as follows:

> Basis: an atomic formula is "translated" by prefixing a necessity operator to it. (In our earlier fantasy about property development, this corresponds to the assumption that the developer won't pull down already completed buildings.)
>
> Recursion:
>
> (i) the translation of a conjunction, disjunction or existential quantification is simply the conjunction, disjunction or existential quantification of the translations of its conjuncts, disjuncts or matrix
>
> (ii) the translation of a negation, conditional or universal quantification is obtained by prefixing a necessity operator to the negation, conditional, or universal quantification of the translation(s) of the formula negated, the antecedent and consequent, or the matrix.

Exactly the theorems of intuitionistic first-order logic translate into theorems of a version of first-order S4 incorporating the Converse Barcan Formula but not the Barcan Formula.

This is formally all very nice (and considering modal translations allows easy proofs of many metamathematical theorems about intuitionistic logic), but what does it mean? S4 has often been proposed as an *epistemic* logic, with the necessity operator interpreted as expressing knowability (Hintikka), or (as in Gödel's abstract) provability. Around 1980 several authors (notably Goodman and Shapiro) suggested a program of "epistemic mathematics": the use of an explicit modal operator within classical mathematics to recover, in a classical context, the useful distinctions (between, for example, constructively proven existence theorems and mere "nonuniversality" theorems) that intuitionism displays. One formal result came out of this work: the exact translation of the intuitionistic into modal systems extends beyond logic to first order Arithmetic.

At a conceptual level, epistemic mathematics doesn't seem to capture the *critical* side of constructivism. (To be fair, its proponents didn't claim it did.) Brouwer and other constructivists rejected some classical theorems, but "epistemic arithmetic," for example, is a proper extension of classical PA. In the case of arithmetic there is an additional potentially embarrassing feature. One trivial theorem of classical arithmetic which is asserted (and asserted to be necessary) in epistemic arithmetic is that every natural number has a successor. As a consequence, every possible world in a model of epistemic arithmetic, or of a Kripke model of Heyting Arithmetic, must have infinitely many objects in its domain, including objects corresponding to every natural number! The idea that the natural number series is "under construction" seems to be represented in these models only by allowing some worlds to contain, in addition to all the "genuine" natural numbers, further non-standard numbers.

This can, however, be avoided. Intuitionistic logic and arithmetic can be interpreted in modal systems with modal logics weaker than S4, and in particular modal logics lacking the principle that all "necessary" propositions are true. Such logics are often interpreted deontically (necessary = it ought to be the case that) or doxastically (necessary = the subject's beliefs logically commit them to believing that), the latter, at least, suggesting an interpretation of intuitionism as a sort of fictionalism about mathematical objects. For a variant of our fantasy about ethical property developers, the necessity

operator could perhaps be interpreted as something like "it is the case that, or, if not, the development plan we are committed to following will bring it about that." A suitable logic of this sort can be obtained from S4 by replacing the principle

If Nec(A) then A

with the weaker

If NecNec(A) then Nec(A);

an interpretation of intutitionistic logic in this logic will use a modified translation scheme on which the translations of disjunctions and existential quantifications, as well as negations, conditionals and universal quantifications, start with a necessity operator (for more detail see Hazen 1990). This allows an interpretation of intuitionistic arithmetic in which every "possible world" has a finite domain. (The interpretation stands to E.W. Beth's semantics for intutionistic logic roughly as the translation into S4 does to Kripke's semantics.)

5

Model-theoretic accounts of intuitionistic systems are typically studied with the methods of classical mathematics; they are, as it were, a nonintuitionist's way of explaining intuitionism. They have, however, been studied constructively, and there are intuitionistic proofs of completeness theorems for parts of intuitionistic logic.

There have also been nonmodel-theoretic semantic accounts of intuitionistic logic and particular intuitionistic mathematical systems, notably S.C. Kleene's notion of *realizability*, and the related theory of constructions of N.D. Goodman. It is surprising that these have not drawn more attention from philosophers of language, for they can be seen as the most precise and detailed attempt to work out what, in the metaphysical literature, is called a theory of *truthmakers*. The idea here is that, for each true proposition, there is an object whose existence *makes* the proposition true. In the metaphysical literature, truthmakers have often been taken to be facts or states of affairs (cf. Olson), or in some cases ordinary objects. In Kleene's original version or realizability, the *realizer* of a true arithmètic statement is a number, taken as encoding (*via* a scheme of Gödel-numbering) some complex of numbers and functions. Thus, for example, the realizer of a true existential quantification encodes some particular witnessing number together with a realizer for the statement that this number realizes the matrix. The realizer of a statement of the form "for every m there exists an n such that . . ." will encode an algorithm computing, for an input m, an appropriate n; the realizer of a statement of the form "for all n, either F(n) or Not-F(n)" will encode an algorithm for deciding, with respect to any given n, whether it is F or not, and so on. The notion (and variations on it) has proven technically fruitful: for example, since all theorems of Heyting Arithmetic have realizers, one way of showing that a sentence is not a theorem is to show that it lacks realizers.

As the multiplicity of interpretations suggests, intuitionistic logic is a a very natural logic with nice properties. Most people familiar with it have a difficult to formulate but intuitively vivid sense that the proof theory of intuitionistic logic, and in particular its natural deduction formulation, is "cleaner" than that of classical logic. It has turned out to have many applications. Some of these are obviously related to its constructivist motivations: computer scientists have taken an interest in intuitionistic logic and

mathematical systems formulated in it because potentially useful algorithms can often be extracted from a proof in a constructive system, as suggested by the examples of the information encoded by realizers in the previous paragraph. On the other hand, intuitionistic logic turns out to be the "correct" one for formalizing domains, such as Smooth Infinitesimal Analysis, which are not obviously inspired by the constructivistic critique of classical mathematics.

It should be noted, however, that adoption of an intuitionistic or constructivistic standpoint does not necessarily commit one to this particular formalization of propositional and first-order logic. Intuitionism is a doctrine about the content of mathematics; the language in which this content is expressed is of secondary interest. If we define our logical constants differently, we can express the same content with a different formal logic. In standard intuitionistic logic, a disjunction can only be proved if one of the disjuncts is provable, but the negation of a conjunction can be established by a simple reductio. One could define a "constructive" kind of negation, for which reductio fails and the negation of a conjunction is provable only if the negation of a conjunct is. Kleene's student David Nelson developed a system of this sort: the logic of constructible falsity. In compensation for abandoning reductio, this logic incorporates many familiar principles: the double negation equivalence, all the De Morgan equivalences (some fail in Heyting's logic), and the analogues of De Morgan's laws for quantifiers. Nelson showed that this logic could be used to express the same mathematical content as standard intutitionistic logic: systems of arithmetic based on Heyting's and Nelson's logics are intertranslatable.

6

Intuitionism is not just about logic, and intuitionistic mathematics is not simply the result of restricting the logical inferences allowed in classical mathematics. Even when they assert (verbally) identical theorems, classical and constructive mathematicians often use different proofs. Because they don't have the cheat of indirect proof available to them, constructivists have been forced to develop methods which, though classically valid, were not discovered by classical mathematicians. There is a deeper kind of difference, however; Brouwer and his successors defined new concepts that have no role in classical mathematics.

Brouwer's theory of real numbers is an example. Real numbers are a generalization of fractions: not all reals are rational. We appeal to them in giving mathematical treatments of measurable quantities, where measurement is conceived as susceptible of ever greater precision: a real number, then, is conceived as the limit of a potentially infinite sequence of rational approximations. With appropriate codings, real numbers can be thought of as given by infinite sequences of natural numbers, but an infinite sequence is not something an intuitionist will believe in as a completed whole. One option is to treat talk of infinite sequences as a *façon de parler* for algorithms generating them, and interesting mathematical theories of real numbers can be developed along these lines. (This is the approach adopted by the school of constructive mathematics stemming from the worrk of the American mathematician Errett Bishop.) Brouwer chose instead a less constrained notion, that of a *choice sequence*: a sequence of numbers determined by the successive choices of a free agent. (To ensure that the sequence is infinite, we can imagine some convention for continuing it if the chooser dies or loses interest, but this is not a major concern. Intuitionism is interested in mathematically possible future

developments, and there is no mathematical guarantee that the chosing won't go on forever.) At any given time, only a finite initial part of the sequence will have been chosen: it is still "under construction," and it cannot be said to have or lack properties that depend on choices that haven't yet been made. In particular, suppose two people are choosing numbers (and so, given some recipe for converting infinite sequences of naturals into real numbers, determining real numbers), and have so far made the same choices. There is no mathematical guarantee that they will continue to make the same choices, and so no guarantee that the real numbers they are generating are the same, but also no guarantee that their choices will diverge. The resulting theory of real numbers differs dramatically from that in classical mathematics (though it does yield most of the standard theorems needed for applications in science and engineering). One consequence is of logical interest. In propositional logic, any classically contradictory formula is also intuitionistically contradictory (and the double negation of any classical tautology therefore intuitionistically valid), but this does not extend to first order logic: the negation of a valid formula of first order logic is not in general intuitionistically contradictory. By appeal to the theory of choice sequences, and the theory of real numbers based on it, we can show that this is not a mere empty possibility. In fact it is a theorem of the intuitionistic theory of real numbers that it is not the case that, for every pair of real numbers x and y, either $x = y$ or it is not the case that $x = y$.

Another example, suggested by Brouwer later in his career and not accepted by all intuitionists, is the theory of the Creative Subject. This is an attempt to make use, within mathematics, of the idea that a proposition is only true if proven. The idea is to consider a special kind of choice sequence, chosen by the mathematician in accordance with the rule "choose 1 if, at the time of choice, you have proven proposition A, but choose 0 if you haven't." If truth is the same thing as proof, then it is perhaps plausible to say that the proposition is true if and only if this sequence contains a 1. Eliminating the reference to the choosing subject, Saul Kripke formulated this as an axiom schema, Kripke's Schema, that could be incorporated in an intutionistic axiom system having variables both for natural numbers and for choice sequences:

> There exists a sequence s such that A if and only if for some n, the n-th number in s is 1.

This can be used to obtain the power of higher-order logic: let sequences stand for sets, and use Kripke's Schema to give the effect of the comprehension schema.

References

Brouwer, L.E.J. (1908). "Over de Onbetrouwbaarheid der logische Principes" ("On the Untrustworthiness of the Principles of Logic"), *Tijdschrift voor Wijsbegeerte*, vol. 2.

Hazen, Allen. (1990). "Actuality and Quantification", *Notre Dame Journal of Formal Logic* 31(4):498–508

5.9
RICHARD MONTAGUE'S APPROACH TO THE SEMANTICS OF NATURAL LANGUAGES

Richmond H. Thomason

1 Logic and Psychology in Semantics

In the nineteenth century, the foundations of logic were a matter of debate. In [Frege, 1953], Gottlob Frege ably argued that logic should be assigned to mathematics rather than to psychology. The conception of logic as a branch of mathematics has dominated the field since that time. Logic is generally considered to be a part of mathematics, and the research methods that logicians use are mathematical.

Similar issues reappear in the twentieth century, in connection with cognitive science. In the 1950s, semantics began to appear as a respectable part of linguistics, but the foundational questions this raises—is it mathematics or psychology?—are less clearly and uniformly resolved. Linguistics is heterogeneous, drawing on psychological and sociological methods, experimentation, data analysis, and applied mathematics. Probably most linguists working in syntax and semantics follow Chomsky in believing that their subject matter is psychological. But, although there are alternatives to formal semantics, such as "cognitive semantics" ([Jackendoff, 1996, Talmy, 2000a]), the prevailing approach to linguistic semantics, based on Richard Montague's work, is thoroughly mathematical. Although some people, notably Barbara Partee ([Partee, 1978, Partee, 1979]) have addressed the discrepancy between the practices and foundational beliefs of semanticists, the tension has not been much addressed. Probably the main reason for the acceptance of logical techniques in semantics is that there is really no competitor when it comes to articulating a detailed, explanatory theory.

So the predominant approach to linguistic semantics, deriving from Montague's work, is based on mathematical logic. This has several consequences.

2 Formalization and Formalized Languages

The systems of logic developed by the early logicists (and especially, Frege and Russell) were concerned prinarily with explicating the notion of a correct mathematical proof,

and were not equipped with a semantics. Since then the techniques for formalization have deepened and sharpened, and apply to a much broader spectrum of languages, with many uses.

Formalizing a language begins with the presentation of an explicit syntax. In the simplest case, this is simply a definition of the formulas of the language. In more general cases, FORMULA or SENTENCE is just one of many *syntactic categories*, and the syntax must define the set of well-formed expressions belonging to each of these categories. Depending on the purpose of the formalized language, the formalization may include a characterization of the associated proofs, or it may articulate some associated algorithms. (Proofs are not important for most programming languages, but a compiler is essential.)

Finally, a formalized language needs a semantic component. Without a semantics, a formalization is blind, lacking a criterion of correctness for proofs and algorithms. In the form developed by Alfred Tarski and used by Montague, a semantics has two parts: (1) a characterization of the *models*, the structures that serve to interpret the language, and (2) a definition of the satisfaction relation, according to which every formula of the language takes on a truth value. Where M is a model and ϕ is a formula, $[\phi]_M$ = T if ϕ is true in M, and $[\phi]_M$ = F if ϕ is false in M.

3 Natural Languages and Formalized Languages

Montague opens his article [Montague, 1970a] with the following noteworthy claim.

> I reject the contention that an important theoretical difference exists between formal and natural languages.

As far as I can tell, the rejected contention never existed in the form of a well-developed and defended position. Certainly it was accepted at the time by many philosophers—ordinary language philosophers as well as their formally inclined colleagues—that the artificial languages studied by logicians and natural languages like English were different in kind. And logic textbooks by thoughtful logicians, such as [Quine, 1940] and [Church, 1959], begin with informal accounts of differences between natural languages and the "well constructed" languages used by logicians, with particular emphasis on the ambiguities to be found in natural languages. The differences these authors note are important, but they don't claim they are in any sense "theoretical."

The idea that the techniques used by logicians could be applied to natural languages, and that natural languages are in principle formalizable along similar lines, was not new. It is mentioned (briefly, but with approval) in [Carnap, 1934] and [Tarski, 1944]. Yehoshua Bar-Hillel took up Carnap's idea with enthusiasm in [Bar-Hillel, 1963]. Bar-Hillel was more familiar with linguistics than most philosophers of his time. Citing work by structural linguists as well as early transformational work by Zelig Harris and Noam Chomsky, he presses the idea that no differences in principle exist between natural languages and the constructed languages studied by logicians, and suggests that logical research could be useful in linguistics, especially in semantics. In a review of [Bar-Hillel, 1954], an earlier paper along simpler lines, Noam Chomsky rejects the proposal; see [Chomsky, 1955]. Chomsky's review, which argues in effect that the idealizations made by logicians render their work largely irrelevant to linguistics, is the most explicit expression of which I'm aware of the view that natural languages are different in

principle from logical languages. I don't know if Montague was familiar with this review. He certainly was aware in general terms of Chomsky's work: In "English as a Formal Language" he says that he that he doesn't consider the work of Chomsky and his students to be helpful in advancing the goal of "serious syntax and semantics."

The ideological contrast here is rather stark, and despite the efforts of people like Barbara Partee, I don't believe it has ever been successfully reconciled. It is therefore somewhat surprising that the main influence of Montague's work has been within linguistics.

In any case, what was dramatically new about Montague's contribution was not the opening salvo, but the way he turned his claim into an explicit research project, which he then proceeded to carry out. After a brief preamble in [Montague, 1970a], he immediately begins to develop an explicit syntax for an extended and ambitious "fragment" of English. Montague preferred to avoid syntactic complications by restricting himself to constructions that—although they are definitely grammatical English—sound stilted and unnatural. For instance, his fragment contains *Every man such that that man loves a woman in Amsterdam walks rapidly*, rather than the more natural *Every man who loves a woman in Amsterdam walks rapidly*. Nevertheless, the sheer scope of this fragment and the others that Montague presented at about this time is astonishing.

4 Montague's Work in Semantics

Montague's publications bearing directly on natural language semantics began in 1968, and his work in this area continued until his death early in 1971. The earlier papers, [Montague, 1968] and [Montague, 1970b], were concerned with developing Intensional Logic—the logical part of the semantic project. Three later papers, [Montague, 1970c], [Montague, 1970a], and [Montague, 1973], develop and describe the general framework, and present grammars for "fragments" of English, treating these fragments using the logical methods described in Section 2, with an explicit syntax and a model-theoretic semantics. A sixth paper published during this period, [Montague, 1969], is an outlier. This paper describes Intensional Logic and advocates it as a framework for doing rigorous philosophy. These papers, and some others, were published together in the posthumous collection [Montague, 1974].

As far as I know, no history of Montague's intellectual development is available, and materials may not exist for such a work, unless someone interviews those still alive who knew him during this period. Montague was a logician first and foremost, but unlike most mathmatical logicians of his time, he believed that the rigorous mathematical techniques of logic could be used in areas other than mathematics, including philosophy. According to [Partee, 2009]—drawing on a personal communication from Hans Kamp—Montague was frustrated, in teaching introductory logic, by the informality and lack of rigor in the part of the course dealing with the relation between logical formulas and English, and this was a large part of the motivation for his work in the semantics of natural language.

David K. Lewis taught at UCLA from 1966 through 1969, and according to [Partee, 2005] Montague and Lewis talked often during that period. Lewis's approach to formal semantics, published in [Lewis, 1970], is similar to Montague's in many ways; it differs in lacking the syntax and semantics for an ambitious "fragment," but compensates for this by presenting ideas for dealing with sorts of constructions that Montague never mentions. Clearly, the two must have influenced each other in many ways.

5 Words, Phrases, and their Meanings

When we seek a meaning, it is usually the meaning of a word. So it seems that once we understand the meanings of a sentence's words, the meaning of a sentence or any other grammatical phrase comes for free. In the leading linguistic theories of meaning, and certainly in the theories derived from logical approaches, this is reversed. The theory has to account for the meanings of phrases, including sentences. This leads theoreticians to concentrate on the linguistic analogues of satisfaction definitions in logic. The task of providing *semantic* rules, showing how the interpretations of grammatical phrases depend on the meanings of their parts is sufficiently complicated to attract most of the attention. Words must have interpretations that are compatible with the semantic rules, but often this means only that their interpretations must have suitable types. The specific interpretation of a lexical item turns out to be of little importance. This is fortunate for the theories, because in most cases it would be difficult or impossible to provide appropriately specific interpretations. Typically, the only natural language words that receive specific interpretations are those, like *every*, *and*, and *or*, that correspond to "logical constants."

On the other hand, natural languages contain many more syntactic constructions and ways of combining them than typical artificial languages. A sentence like

(1) *Every decent man who loves a woman who loves him should eventually want to marry her*

is relatively simple and perfectly intelligible, but at least nine different constructions interact in producing the interpretation. (1) The combination of a transitive verb with its direct object, as in *loves+a woman*. (2) The combination of a verbal phrase with its subject, as in *he+loves a woman*. (3) The combination of a relative clause with a nominal phrase, as in *woman+who loves him*. (4) The combination of an article or quantifier with a nominal phrase, as in *a+woman*. (5) The combination of a tense with a verbal phrase, as in PRESENT+*love=loves*. (6) The combination of a modal auxiliary with a verbal phrase, as in *should +eventually want to marry her*. (7) The combination of an adjective with a nominal as in *decent+man*. (8) The combination of an adverb with a verbal phrase, as in *eventually +want to marry her*. (9) The combination of a verb with an infinitive phrase, as in *want +to marry her*. (In these examples, 'combination' means 'syntactic combination', and '+' denotes a syntactic operation. For example, 'nominal phrase' and 'relative clause' are to be understood as syntactic categories, and the word order that results when they combine would be given by a syntactic rule. But in these examples, I am being informal and somewhat arbitrary about the syntax.)

It would be impossible to account for the meanings of even moderately complex sentences without an architecture for classifying possible meanings and their interactions. Perhaps Montague's most important new contribution was to provide such an architecture, and to show how it could be systematically used to explain the meanings of natural language phrases and their modes of combination. It is appropriate that Montague's publications on natural language semantics begin with Intensional Logic, because the materials for producing appropriate meanings lie in the logic.

6 A Logic Based on Functions

The idea that meanings combine with other meanings to form new meanings creates a fundamental need for suitable modes of combination. Montague looks to logic to provide these modes, and logic prizes parsimony: Many of the most impressive results in logic show how to produce remarkable expressive results from minimal resources. The best theory would manage with only one mode of combination. The most likely candidate for this one mode is functional application. Functions are pervasive in mathematics, and mathematics provides powerful mechanisms for defining functions.

Moreover, functions are plausible vehicles for combining the meanings of constituents into larger meanings. For instance, it makes a certain amount of sense to say that the meaning of an adjective like *decent* is a function taking the meaning of a nominal phase, such as *man*, into the meaning of a new nominal phrase, *decent man*. If the interpretation of *decent* is a function f, then [[*decent man*]], the interpretation of the phrase *decent man*, will be f([[*man*]]): so here we have obtained the meaning of the phrase *X + Y* by taking the meaning of *X* to be a function, and applying this function to the meaning of *X*.

If the composition of meanings is a matter of functional application, the logic of meaning composition will be a logic of functions. At this point, earlier developments in logic dovetail nicely with the needs of semantics. These developments provide not only an appropriate logic of functions but the crucial idea of a *logical type*.

Both Frege and Bertrand Russell undertook the project of producing a logic that would allow mathematics, and especially the mathematics underlying the calculus, to be presented by means of a series of definitions. Frege's combination of set theory and logic led to a formalization that Russell showed to be inconsistent, replicating the paradox of the set consisting of all sets that are not members of themselves. Russell's formalization of the mathematics of real numbers, in [Whitehead and Russell, 1910–1913], avoided this paradox by enforcing a system of *logical types*; a set must have a type higher than the type of any of its members.

After the publication of *Principia Mathematica* in 1910, a series of developments in logic produced simplified and improved formalizations of the theory of logical types. These developments culminated in 1940, with Alonzo Church's presentation in [Church, 1940] of a version of type theory based entirely on functions. Functions and sets are theoretically interchangeable. On the one hand, you can regard a function as a set of ordered pairs, and on the other, you can identify a set X with the function f_X that outputs the truth-value T for the input x if $x \in X$, and otherwise outputs F.

Combining the notion that every object in the ontology of the logic has a type with a logic based on functions, we arrive at the idea that two types must be associated every function: (1) the type of the function's domain, of its inputs, and (2) the type of the function's range, of its outputs. A function from individuals to truth-values, for instance, would be characterized by two components: the type of individuals and the type of truth-values. These last two types are not functional; they are basic.

The system of all types can then be characterized as follows: (1) there are two basic types, the type e of individuals and the type t of truth-values; (2) if σ and τ are types, then $\langle \sigma, \tau \rangle$ is the type of functions from things of type σ to things of type τ; (3) there are no other types. For instance, $\langle e, t \rangle$ is the type of functions from individuals to truth-values, and $\langle e, \langle e, t \rangle \rangle$is the type of functions from individuals to functions from individuals to truth-values.

Church's language for the logic of functional types has a family of equality symbols—for each type τ there is an equality constant of type $\langle\tau, \langle\tau, t\rangle\rangle$—and relies on only two constructions: functional application and lambda abstraction. Functional application involves an expression ζ of type $\langle\sigma, \tau\rangle$ and an expression η of type σ; $\zeta(\eta)$ is then an expression of type τ, and denotes the result of applying the function denoted by ζ to the object denoted by η.

Lambda abstraction is more complicated, but it proved to be highly useful in Montague's linguistic applications of type theory. It is a general method for constructing explicit names of functions. For instance, in arithmetic, $\lambda x\, x^2 + 1$ denotes the function that inputs a number and outputs the result of squaring that number and adding 1. In general, an expression ζ of type τ, containing a variable, say x, of type σ, will denote an implicit function whose output is the denotation of σ when x is assigned an input denotation of type σ. $\lambda x\zeta$ explicitly denotes this function, which has type $\langle\sigma, \tau\rangle$. Lambda abstraction on formulas of type t is a special case: $\lambda x^e[x = x]$, for instance, is the constant function that inputs an individual and outputs the truth-value t.

7 Typing Phrases of Natural Language

The apparatus of typing, originally motivated by the paradoxes, has become useful in a remarkably broad range of applications. If Church's types can be assigned in a plausible way to the phrases of natural language, this will support a method of semantic interpretation that uses functional application to interpret syntactic combinations. Montague showed that this can be done, and the methods he developed continue to be used by formal semanticists.

For instance, take the sentence

(2) *Andy is single and unemployed.*

The subject, *Andy*, is a proper name; if we assume that proper names denote individuals, then *Andy* will have type e. We can leave the verb *is* uninterpreted or assign it the identity function. The phrase *single and unemployed* is what logicians would call a one-place predicate, so it would have type $\langle e, t\rangle$. These types allow us to obtain the interpretation of the entire sentence by functional application, applying the denotation of *single and unemployed* to the denotation of *Andy*. The adjectives *single* and *unemployed* have type $\langle e, t\rangle$. Therefore, the *and* that combines them must have type $\langle\langle e, t\rangle, \langle\langle e, t\rangle, \langle e, t\rangle\rangle\rangle$. This last type indicates a function that inputs (sequentially) two adjective denotations, and outputs an adjective denotation. These type assignments make good intuitive sense, and provide a functional interpretation of the entire phrase.

Of course, there will be an *and* that combines sentences, as in

(3) *Andy is single and Andy is unemployed.*

This *and* will have type $\langle t, \langle t, t\rangle\rangle$. Since each lexical item can have only one logical type, we have to declare (at least) two senses of *and*: an and_1 that combines sentences and an and_2 that combines adjectives.

8 Interpretation, Translation, and Meaning Postulates

Once the syntax of a fragment has been specified, its semantics is given by a model-theoretic interpretation. The models of the theory are specified, and for each model M, the interpretation $[\zeta, D]_M$ of the phrase ζ in M has to be defined for each phrase ζ of the fragment, where D is a disambiguator. (Disambiguators will be explained in Section 9; for the moment, I'll ignore this complication.)

The model-theoretic interpretation can be done in either of two ways. The interpretation can be direct, with a satisfaction definition like those that are used for logical languages. Or it can be indirect, using a logical language that already has a model-theoretic interpretation and defining a translation into the logic for each disambiguated phrase. Montague sometimes (for instance, in [Montague, 1970a]) uses the direct method, and sometimes (in [Montague, 1973]) the indirect method.

The indirect method is preferable, because it provides a logical language (the target of the translation) in which to state *meaning postulates*. The need for such postulates in semantics was somewhat of an embarrassment for Carnap, but they play a different and important role in Montague's theory, making it possible to nullify the complicating effects of the apparatus of types. We can illustrate this point with the two *and*s of Section 7.

Since and_1 is a logical constant, we can completely specify its meaning by means of a meaning postulate. In Postulate (4), and_1' is the logical translation of and_1. (This follows Montague's notational practice in [Montague, 1973].) Using lambdas, the postulate says that this sense of *and* means the same thing as the logical conjunction $\wedge$ (which is available in Church's logic).

(4) $and_1' = \lambda x^{t} \lambda y^{t} [x \wedge y]$

Another postulate relates the interpretation of the two *and*s.

(5) $and_2' = \lambda x^{\langle e,t\rangle} \lambda y^{\langle e,t\rangle} \lambda z^{e}[x(z) \wedge y(z)]$

It follows from this postulate that Examples (2) and (3) are logically equivalent. With such equivalences, the need to proliferate *and*s is not so harmful.

Montague's uniform treatment of noun phrases as generalized quantifiers uses similar techniques, and was especially influential. Proper names like 'Andy' and quantificational noun phrases are similar in their syntactic distribution and behavior; linguists classify them together as *noun phrases* or NPs.

Quantificational phrases have a natural place in Church's system of types. Since Frege introduced the idea, logicians have been familiar with the identification of numbers as sets of sets: *five dimes*, for instance, would correspond to the set of all sets containing five dimes. In functional terms, this would mean that *five* has type $\langle\langle e, t\rangle, \langle\langle e, t\rangle, t\rangle\rangle$, so that *five dimes* will have the type $\langle\langle e, t\rangle, t\rangle$. But other quantificational phrases, such as *every dime* and *a dime* can be interpreted in the same way; the former corresponds to the set of sets that contain all dimes, and the latter to the set of sets that contain at least one dime. Using meaning postulates, this would mean, for instance, that the translation of *every* would satisfy the following constraint.

(6) every $= \lambda x^{\langle e,t\rangle} \lambda y^{\langle e,t\rangle} \forall z^{e} [x(z) \rightarrow y(z)]$

With this new type assignment to NPs, the interpretation of the subject–predicate relation remains functional but is reversed: in a sentence like

(7) *Every dime is round*,

round has type ⟨e, t⟩, *every dime* has type ⟨⟨e, t⟩, t⟩⟩, and the meaning of the whole sentence is obtained by applying the function that the subject denotes to the set denoted by the predicate. In effect, the sentence is interpreted as saying "The set of round things is a member of the set of all sets containing every dime." This policy towards NPs, together with the uniform use of function application in translation, makes for logical translations that look more like English than logical formulas. For instance, the translation of Sentence (7) is Formula (8).

(8) [*every′* (*dime′*)](*round′*)

The textbook translation, $\forall x^{e}$ [*dime′*(x) → *round′*(x)], would be obtained by applying meaning postulates to (8).

Typing constraints in the logic provide a way to verify formulas such as (8), known as *type checking*. The formula *every′* has type ⟨⟨e, t⟩, ⟨⟨e, t⟩, t⟩⟩, giving *every′*(*dime′*) the type ⟨⟨e, t⟩, t⟩. But *dime′* has type ⟨e, t⟩, giving Formula (8) the type t, which is the type assigned to sentences.

By making the interpretation of NPs uniform, we have given proper names—which intuitively denote objects—a higher type than we would expect. Again, meaning postulates can prevent the unwanted effects of this assignment of types. In this case, we want to associate an individual with each proper name, serving as its customary denotation, and invoke a meaning postulate saying that the interpretation of a proper name is the set of sets that contain the associated individual. The following meaning postulate will accomplish this for the name *Andy*.

(9) $\exists x^{e} \forall y^{\langle e, t\rangle}$ [*Andy′*(y) ↔ y(x)]

9 Obstacles to the Formalization of Natural Language

The reasons that led many logicians and philosophers to feel that natural and logical languages were fundamentally different form a list of challenges to Montague's project. These include: ambiguity, vagueness, context-sensitivity, and intensionality. It is interesting to see how Montague addressed them in carrying out his project.

Ambiguity. Montague, along with many subsequent semanticists, assumes that the only sources of ambiguity are lexical and structural. Lexical ambiguity is familiar; it arises whenever a single word has two or more different senses. Structural ambiguity occurs when the syntactic rules can produce the same phrase in more than one way. The sentence *I know what you know*, for instance, is structurally ambiguous, since *what you know* can either be an embedded question or a relative clause—and it has two different meanings, corresponding to the syntactic alternatives.

A disambiguator for a phrase, then, is a specification of a sense for each lexical item in the phrase and of a syntactic derivation for the phrase. The interpretation of a phrase must be relativized to a disambiguator.

If structural ambiguities can't be explained without finding appropriate syntactic derivations, semantic evidence about ambiguity can precipitate decisions about how

the syntax should be arranged. This happens in Montague's fragments in the case of sentences that exhibit quantifier-scope ambiguities, such as Montague's example, *Every man loves a woman*. Conservative syntacticians at the time found the idea of semantic arguments for syntactic structure unwelcome, but now such treatments of quantifier-scope ambiguities are widely accepted in linguistics. Montague's approach to semantic ambiguity seems, in fact, to have held up well.

Vagueness. Montague offers no solution to the problem of vagueness. This topic has been studied extensively—mostly in the philosophical literature—since 1973. In my opinion, we do not have an entirely successful solution to the problem yet. However, this doesn't seem to be problem specifically for the semantics of natural languages, but for any language intended to be used in domains where complete precision is lacking.

Context-sensitivity. Montague considered this problem in his papers on pragmatics. His solution, which occurred independently to many others, was to make interpretation in a model sensitive to various contextual factors. An extensive subsequent literature developed from this starting point. This is an area where the theories have changed and gradually become more sophisticated. But these developments enhance Montague's program of formalizing natural languages, rather than threatening it.

Intensionality. The problem of intensionality was, of course, stressed by Willard V.O. Quine and other philosophical skeptics about theories of meaning; see, for instance, [Quine, 1953]. It is a direct and glaring problem for the scheme of interpretation as we have presented it so far.

In interpreting natural language, the problem of intensionality centers on the logical type that is associated with sentences. A propositional attitude construction, such as *believe*, combines with a sentence to form a verb phrase: The present tense form of *believe*, for instance, combines with *Los Angeles is the capital of California* to form the VP *believes that Los Angeles is the capital of California*. This then combines with a subject like *Bertha* to form the sentence

(10) *Bertha believes that Los Angeles is the capital of California*.

So the interpretation [[*Bertha believes that Los Angeles is the capital of California*]] of the entire sentence (which is just a truth-value, T or F, in Church's type theory) will be a function of the interpretation [[*Los Angeles is the capital of California*]] of the subordinate clause (which is also a truth-value). Something has gone very wrong, since this implies that if Example (10) is true, then Bertha believes any false proposition whatsoever. Montague addresses the problem by providing an enhanced interpretation of sentences, based on possible-worlds semantics: the new interpretation is a set of possible worlds, or—to put it in functional terms—a function from possible worlds to truth-values. Possible-worlds semantics—and especially, the theory of how quantifiers interact with modal operators—was relatively new and undeveloped as a branch of logic when Montague turned his attention to it. Montague provided one solution to the problem of how to combine quantification with modality, while retaining a reasonable model-theoretic interpretation. But he went much further than this, extending modal logic to include higher-order quantification and generalizing earlier ideas from [Carnap, 1956]. Montague called the logical system that resulted from these ideas "Intensional Logic." IL provides an alternative to Frege's theory of sense and denotation with a well-developed model-theoretic interpretation.

10 Intensional Logic

Although Montague didn't formulate it quite this way, it is easiest to think of Intensional Logic as the type theory that results from the addition of a new type, w, to Church's type theory—the type of possible worlds. Call the *extensional types* of the new logic those that do not involve this new type. For each extensional type τ (and indeed for any type τ), there is a corresponding *intensional type*, $\langle w,\tau\rangle$.

Only two changes are needed in the models and satisfaction definition for the extended theory. (1) Satisfaction is relativized to a world: $[[\zeta]]_{M,w}$ is the value assigned to ζ in M at world w. (2) Two new constructs, $^\wedge$ and $^\vee$, need to be interpreted. The "up" operator, $^\wedge$, is an intensionalizer; if ζ is a formula of type τ, $^\wedge\zeta$ is a formula of type $\langle w,\tau\rangle$. In effect, $^\wedge$ performs lambda abstraction over an implicit world variable: where ζ has type τ, $[[^\wedge\zeta]]_{M,w}$ is the function f from the set of worlds to objects of type τ such that $f(w) = [[\zeta]]_{M,w}$. The $^\vee$ operator is an extensionalizer of formulas of intensional type. If ζ has type $\langle w,\tau\rangle$, then $^\vee\zeta$ is wellformed and has type τ, and $[[^\vee\zeta]]_{M,w} = [[\zeta]]_{M,w}(w)$.

The intension of the type formerly assigned to sentences is $\langle s, t\rangle$, the type of propositions or functions from worlds to truth-values, and propositional attitudes are then treated as usual in possible-worlds semantics. But there are many other sources of intensionality in natural language, and Montague deserves the credit for recognizing that a type-theoretic modal logic provides the flexibility for dealing quite generally with the phenomena. In courses on Montague Grammar, the point was illustrated with sentences like *John seeks two unicorns*.

Of course, the use of possible-worlds semantics inherits the problems of that approach, especially a commitment to "logical omniscience."

For an exhaustive treatment of IL as a logic, with applications, see [Gallin, 1975].

Linguists have gradually come to accept possible worlds as a working tool, and have found many applications of the idea of which philosophers may not be aware. See [Partee, 1989] for an excellent survey of the topic, as well as the more recent general handbook articles concerning Montague Grammar.

11 From Montague Grammar to Formal Semantics

Barbara Partee reports in [Partee, 2005] (note 36) that she coined the term "Montague Grammar" around 1971. But, as Montague's ideas and later additions have been absorbed into the background of linguistic semantics, and the field has diversified, the term has been replaced by "formal semantics." Linguists have, in general, not taken Montague's approach to syntax very seriously, and so Montague's semantic ideas have been combined with many more popular syntactic frameworks. At the same time, the theories have been fruitfully applied to an enormously wide range of constructions, in English and many other languages. And the logical framework has been enriched, most notably with various dynamic mechanisms. For more information on these developments, see [Partee, 1996] and [Partee and Hendriks, 1996].

Philosophers should take note that the term "Montague Grammar" is somewhat outdated, and that linguists have for some time been pursuing projects in areas of formal semantics that are very relevant to philosophy. There is also much of interest to philosophers in formal pragmatics, a younger field which would not have been possible without relatively mature theories of linguistic semantics. Many papers of interest in both fields have appeared in the journal *Linguistics and Philosophy*, and many others are archived at http://semanticsarchive.net/.

12 Philosophical Consequences

Montague's idea that Intensional Logic provided a way of formalizing philosophy, turning it into a subject much more like mathematics, was unrealistic. And its uncompromising presentation would make it difficult even for sympathetic philosophers to see that there is a genuine and important point here. The philosophical usefulness of a slightly less rigorous approach in the same spirit has been demonstrated in the subsequent work of philosophers like David Lewis and Robert Stalnaker. This influence is already well entrenched in various areas of philosophy, particularly in metaphysics.

Time always works in favor of specialization and division of labor. But there is not yet a happy division between formal semantics/pragmatics and the related areas of philosophy. There is much overlap but too little communication. I believe that it would be beneficial for researchers working in philosophy of language, or developing a point that depends on language, to form a habit of studying the relevant work of formal semanticists. This work can add a new dimension to philosophical project. Linguists are trained to deploy sophisticated theory and systematic use of linguistic evidence in ways that go far beyond what is typical in philosophy. This work is often directly relevant to philosophical issues, and can add new depth to the treatment of these issues.

Looking at language systematically can also be a good way to achieve coverage. To a very large extent, what Emmon Bach calls "natural language metaphysics" in [Bach, 1986]—working out the formal semantics of natural language constructions that seem to incorporate basic ways we conceptualize the world—overlaps with metaphysical projects having to do with topics such as causality, time and change, and mereology. (See the references in [Bach, 1986]. And [Dowty, 1979] is still a useful source on this topic.) But because of the need for linguistic coverage, it is driven by systematic considerations that seem to be missing in the comparable areas of contemporary philosophy.

Relationships between linguistics and philosophy have always been stressed by the journal *Linguistics and Philosophy*. This journal is perhaps the most useful contemporary and historical source of interactions between formal semantics and various areas of philosophy. Some of these interactions are discussed in [Pelletier and Thomason, 2002], a survey editorial that appeared in the twenty-fifth anniversary volume of the journal. And the papers that appeared in the same issue were specially invited to illustrate connections between the two fields.

References

Emmon Bach. Natural language metaphysics. In Ruth Barcan Marcus, Georg J.W. Dorn, and Paul Weingartner, editors, *Logic, Methodology, and Philosophy Of Science, VII: Proceedings of the Seventh International Congress of Logic, Methodology, and Philosophy of Science, Salzburg, 1983*, pages 63–81. North-Holland, Amsterdam, 1986.

Yehoshua Bar-Hillel. Logical syntax and semantics. *Language*, 30(2):230–237, 1954.

Yehoshua Bar-Hillel. Remarks on Carnap's *logical syntax of language*. In Paul A. Schilpp, editor, *The Philosophy of Rudolph Carnap*, pages 519–543. Open Court Publishing Company, LaSalle, Ilinois, 1963.

Rudolph Carnap. *Logische Syntax der Sprache Schriften zur wissenschaftlichen Weltauffassung*. Verlag von Julius Springer, Vienna, 1934.

Rudolph Carnap. *Meaning and Necessity*. Chicago University Press, Chicago, 2nd edition, 1956. (First edition published in 1947.).

Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar: An Introduction to Semantics*. The MIT Press, Cambridge, Massachusetts, 2nd edition, 2000.

Noam Chomsky. Logical syntax and semantics: Their linguistic relevance. *Language*, 31(1):36–45, 1955.

Alonzo Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5(1):56–68, 1940.

Alonzo Church. *Introduction to Mathematical Logic, Vol. 1*. Princeton University Press, Princeton, 1959.

David R. Dowty, Robert Wall, and Stanley Peters. *Introduction to Montague Semantics*. D. Reidel Publishing Co., Dordrecht, 1981.

David R. Dowty. *Word Meaning in Montague Grammar*. D. Reidel Publishing Co., Dordrecht, Holland, 1979.

Gottlob Frege. *The Foundations of Arithmetic*. Oxford University Press, Oxford, 2nd edition, 1953. (Originally published in 1884. Translated by J.L. Austin.).

Daniel Gallin. *Intensional and Higher-Order Logic*. North-Holland Publishing Company, Amsterdam, 1975.

Leon Henkin. Completeness in the theory of types. *Journal of Symbolic logic*, 15:81–91, 1950.

Ray Jackendoff. Semantics and cognition. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 539–559. Blackwell Publishers, Oxford, 1996.

David K. Lewis. General semantics. *Synthése*, 22(1–2):18–67, 1970.

Richard Montague. Pragmatics. In R. Klibansky, editor, *Contemporary Philosophy: A Survey. I*, pages 102–122. La Nuova Italia Editrice, Florence, 1968. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 95–118.

Richard Montague. On the nature of certain philosophical entities. *The Monist*, 53:159–194, 1969.

Richard Montague. English as a formal language. In Bruno Visentini et al., editor, *Linguaggi nella Societá, e nella Tecnica*, pages 189–224. Edizioni di Communità, Milan,1970. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 188–221.

Richard Montague. Pragmatics and intensional logic. *Synthése*, 22:68–94, 1970. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 119–147.

Richard Montague. Universal grammar. *Theoria*, 36:373–398, 1970. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 222–246.

Richard Montague. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius M.E. Moravcsik, and Patrick Suppes, editors, *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, pages 221–242. D. Reidel Publishing Co., Dordrecht, Holland, 1973. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 247–270.

Richard Montague. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven, Connecticut, 1974. Edited, with an introduction, by Richmond H. Thomason.

Barbara Hall Partee. Montague grammar, mental representations, and reality. In Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, editors, *Contemporary Perspectives in the Philosophy of Language*, pages 195–208. University of Minnesota Press, Minneapolis, 1978.

Barbara Partee. Semantics—mathematics or psychology? In Rainer Bäuerle, Urs Egli, and Arnim von Stechow, editors, *Semantics from Different Points of View*, pages 1–14. Springer-Verlag, Berlin, 1979.

Barbara H. Partee. Possible worlds in model-theoretic semantics: A linguistic perspective. In Sture Alén, editor, *Possible Worlds in Humanities, Arts and Sciences*, pages 93–123. Walter de Gruyter, Berlin, 1989.

Barbara H. Partee. The development of formal semantics in linguistic theory. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 11–38. Blackwell Publishers, Oxford, 1996.

Barbara H. Partee. Reflections of a formal semanticist as of Feb 2005. http://people.umass.edu/partee/docs/BHP Essay Feb05.pdf

Barbara H. Partee. Reminiscences on philosophy-linguistics interactions in the beginnings of formal semantics. https://udrive.oit.umass.edu/partee/RichfestPartee.ppt.pdf.

Barbara H. Partee and Herman L.W. Hendriks. Montague grammar. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 5–91. Elsevier Science Publishers, Amsterdam, 1996.

Francis Jeffry Pelletier and Richmond H. Thomason. Twenty-five years of linguistics and philosophy. *Linguistics and Philosophy*, 25(5–6):507–529, 2002.

Willard V.O. Quine. *Methods of Logic*. Holt, Reinhart and Winston, New York, 1940.

Willard V.O. Quine. Three grades of modal involvement. In *Proceedings of the XIth International Congress of Philosophy, Volume 14*, pages 65–81, 1953.

Leonard Talmy. *Toward a Cognitive Semantics: Concept Structuring Systems*, volume 1. The MIT Press, Cambridge, Massachusetts, 2000.

Leonard Talmy. *Toward a Cognitive Semantics: Typology and Process in Concept Structuring*, volume 2. The MIT Press, Cambridge, Massachusetts, 2000.

Alfred Tarski. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4:341–375, 1944.

Richmond H. Thomason. Introduction. In Richmond H. Thomason, editor, *Formal Philosophy: Selected Papers of Richard Montague*, pages 1–69. Yale University Press, New Haven, Connecticut, 1974.

Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*. Cambridge University Press, Cambridge, England, first edition, 1910–1913. Three volumes.

Section VI

PHILOSOPHY OF LANGUAGE FOR THE REST OF PHILOSOPHY

6.1
PHILOSOPHY OF LANGUAGE FOR EPISTEMOLOGY

Ram Neta

The present article surveys some of the ways in which recent epistemology has been informed by work in the philosophy of language. In particular, we focus on four important issues.

1 Is Knowing How to Do Something a Form of Knowing that Something Is the Case?

Gilbert Ryle (1949) begins chapter 2 as follows: "In this chapter I try to show that when we describe people as exercising qualities of mind, we are not referring to occult episodes of which their overt acts and utterances are effects; we are referring to those overt acts and utterances themselves" (Ryle 1949, 25). In the course of trying to show this, Ryle occasionally describes our qualities of mind, or the overt acts and utterances in which we express those qualities, as "knowledge *how*," and he also occasionally describes the occult episodes spuriously posited as the causes of these acts or utterances as "knowledge that." Since the chapter in which these occasional descriptions occur is entitled "Knowing How and Knowing That," it is commonly assumed that Ryle means to be claiming that knowing how to do something is not a matter of knowing that something or other is the case. (In fact, I find nothing in Ryle's text that commits Ryle to this view. What Ryle is committed to is that the know-how that is exhibited in our overt acts and utterances is not a matter of our knowing any rules, and that it is not a matter of our acts and utterances having any particular antecedents.)

But whether or not Ryle thought that knowing how was distinct from knowing that, a number of other philosophers (for example, Putnam 1996, Lewis 1990, Devitt 1996) have assumed that they are distinct. Is this assumption correct? The assumption was challenged by Vendler 1972, who writes:

> It will be objected here that in some cases of knowing how it is impossible to tell, in words, what one knows. I know how to tie a necktie, but I could not tell you in words alone. I grant this, but point out that this situation is possible with nearly all the *knowing wh* forms. I know what coffee tastes like, what the color

> magenta looks like, where it itches on my back, when I should stop drinking, how the coastline of Angola runs, but I could not tell you in words alone. I must have, however, some other means to supplement words: pointing, offering a sample, a sketch, a demonstration; or saying 'now.' By these means I can tell you, or show you, what I know: I know that magenta looks like this (offering a sample), that it itches here (pointing), that I should stop drinking now. The need to supplement words with non-linguistic media affects knowing how, and knowing *wh* in general, exactly because it affects the corresponding knowing that.
>
> (Vendler 1972, 104–5)

The point that Vendler makes in the passage just quoted is echoed by Ginet 1975 and then substantially elaborated and defended by Stanley and Williamson 2001. Stanley and Williamson argue for Vendler's point by appeal to the syntax and semantics of knowledge ascriptions involving embedded questions, for example,

> John knows where the newspaper is.
> James knows why the indicator light is on.
> Tanya knows when the next show begins.
> Sonia knows how to ride a bicycle.

The syntax of such sentences can be represented as follows:

> John knows [where PRO the newspaper is t].
> James knows [why PRO the indicator light is on t].
> Tanya knows [when PRO the next show begins t].
> Sonia knows [how PRO to ride a bicycle t].

PRO is an unpronounced pronoun that occurs in the subject position of the embedded clauses, and t is a trace left by moving the "wh-" term ("where", "why", etc.) into subject position. Syntactically, ascriptions of *knowledge how* are of the same form as ascriptions of *knowledge where*, *knowledge why*, and *knowledge when*. Syntax thus suggests that *knowledge how* is a kind of knowledge *wh-*.

But how should we understand knowledge *wh- . . .*? Since *wh-* clauses express questions, it is perhaps best to start by asking: what is the semantic value of a question? According to the seminal Groenendijk and Stokhof 1984 (which was influenced by Hamblin 1958), the semantic value of a question is the set of propositions that can answer that question, consistently with the presuppositions of the question. For instance, the question "When was Abraham Lincoln born?" has as its semantic value the set of propositions P such that each member of P is a proposition of the form *Abraham Lincoln was born at t*, where t is a time at which, consistently with the presuppositions of the question (for example, that there was exactly one year in which Abraham Lincoln was born), Abraham Lincoln could have been born. And the question "Why don't human beings have wings?" has as its semantic value the set of propositions Q such that each member of Q is a proposition of the form *Human beings don't have wings because s*, where s is a proposition that, consistently with the presuppositions of the question, can explain why human beings don't have wings. In each case, the semantic value of a question is a set of propositions that, consistently with the presuppositions of the question, can constitute an answer the question.

Given this semantics for questions, how should we understand ascriptions of knowledge *wh- . . .*? It is standard to understand such ascriptions as follows: S knows *wh-* = there is a such that p is the true answer to the question expressed by the *wh-* clause, and S knows that p. Applying this generalization straightforwardly to the case of *knowledge how*, we get the somewhat obscure result that S knows how to F = there is a p such that p is the true answer to the question "how to F?", and S knows that p. This result is obscure because it is obscure what sort of question is intended by an interrogative sentence of the form "How to F?" And that, Stanley and Williamson say, is because clauses of the form [how PRO to F t] are ambiguous at two different points. First, the infinitive (for example, "to ride a bicycle") can be used to express a deontic modality, or a nondeontic modality. And second, PRO can refer back to the subject of the unembedded clause (for example, Sonia), or it can be used in a way that is equivalent to the English pronoun "one". Thus, we get four possible interpretations of "Sonia knows how to ride a bicycle." Roughly:

Sonia knows how she herself is able to ride a bicycle.
Sonia knows how she herself ought to ride a bicycle.
Sonia knows how one is able to ride a bicycle.
Sonia knows how one ought to ride a bicycle.

If we interpret "Sonia knows how to ride a bicycle" in the first of the four ways listed above, then we can say this: Sonia knows how to ride a bicycle = there is a p such that p is the answer to the question "How is Sonia able to ride a bicycle" and Sonia knows that p. If we interpret "Sonia knows how to ride a bicycle" in the second of the four ways listed above, then: Sonia knows how to ride a bicycle = there is a p such that p is the answer to the question "How ought Sonia to ride a bicycle?" and Sonia knows that p. If we interpret "Sonia knows how to ride a bicycle" in the third of the four ways, then: Sonia knows how to ride a bicycle = there is a p such that p is the answer to the question "How is one able to ride a bicycle?" and Sonia knows that p. And so on. But notice that, on any of the four interpretations given above, for Sonia to know how to ride a bicycle is for there to be some p (some proposition, or perhaps some fact–see the next section) such that Sonia knows that p. And so, Stanley and Williamson conclude, *knowing how* is a kind of *knowing that*.

2 Is it Propositions that Are Known, or Facts?

Russell 1918 coined the phrase "propositional attitude" to refer to psychological states that have propositional contents: for example, beliefs, wishes, hopes, and so on. To say that these psychological states have *propositional content* is to say that these psychological states have a kind of representational content that is individuated more finely than by its truth-conditions. Most epistemologists assume that knowledge has precisely this same kind of representational content; they assume that what one knows is just the same kind of thing that one believes, wishes, or hopes, and it is a kind of content that is individuated more finely than by its truth-conditions. There seem to be a couple of very simple and decisive arguments for this popular view. First, suppose I ask you: "Do you *believe* that the ship has sailed, or do you *know* it?" It seems that I am asking whether your relation to the proposition that the ship has sailed is the relation of knowledge or of belief. But if it is reasonable for me to ask this question, *so understood*, then it is just

as reasonable for me to suppose that knowledge, like belief, has propositional contents. And so, it is generally thought, knowledge has propositional content, just as belief does. Second, suppose that John knows that Superman is in New York and James knows that Clark Kent is in New York. Superman is identical to Clark Kent, but does it follow that John and James know the very same thing? It seems that this does not follow, and so it may seem that the content of knowledge, like the content of belief, is individuated more finely than by its truth-conditions. Again, it is concluded, knowledge has propositional content, just as belief does.

Vendler 1972 challenged this orthodoxy, and the arguments for it. Vendler adduced evidence for the hypothesis that, while belief is a relation to a proposition, knowledge is a relation to something that is not a proposition. Knowledge, on Vendler's view, is a relation to a fact. What is the difference between a proposition and a fact? And why does Vendler think that beliefs are relations to the former and knowledge to the latter? Let's address these questions in order.

The difference between a proposition and a fact is indicated by a couple of things. First, some propositions are true and others are false, but facts are neither true nor false. Facts, unlike propositions, do not have truth-values. Second, substitution of co-referring terms in a clause need not preserve which proposition is denoted by that clause. But if there is a fact that corresponds to that clause, then the very same fact corresponds to any clause that results from substitution of co-referring terms in a referential transparent context. For instance, the proposition that Hesperus is in the sky differs from the proposition that Phosphorus is in the sky: those who are ignorant of astronomy might reasonably believe one of these propositions without believing the other. But the fact (if it is a fact) that Hesperus is in the sky is none other than the fact that Phosphorus is in the sky.

Now that we've offered these two indications of the difference between facts and propositions, let's consider Vendler's argument for the claim that we *believe* propositions, but we *know* facts. In support of this claim, Vendler adduces some evidence.

First, consider the following sentences:

(a) John believes that grass is green.
(b) James believes what John believes.
(c) James knows what John believes.

(d) John suspects that grass is green.
(e) James believes what John suspects.
(f) James knows what John suspects.

The conjunction of (a) and (b) implies that James believes that grass is green, and so does the conjunction of (d) and (e). But the conjunction of (a) and (c) does not imply this, nor does the conjunction of (d) and (f). But, if the only difference between (b) and (c), and between (e) and (f), is the substitution of "knows" for "believes," then why should there be this difference?

Second, notice the difference between:

(g) James knows where John's car is parked.
(h) * James believes where John's car is parked.

(i) James knows why John parked his car there.
(j) * James believes why John parked his car there.

(k) James knows how to park a car.
(l) * James believes how to park a car.

The first sentence of each of the above pairs is perfectly grammatical, but not the second is clearly ungrammatical. Why?

Here's how Vendler answers these questions. Knowledge, for Vendler, is a relation that we bear to facts, and facts, unlike propositions, are denoted by *wh-* clauses, for example, "whom Onassis married", "when the party will start", "why the isotope decayed", and so on. (Notice that *wh-* clauses do not have truth-values, and are referentially transparent.) That's why (g), (i), and (k) are grammatical. But belief is a relation that we bear to propositions, not to facts. That's why (h), (j), and (l) are ungrammatical.

But then if belief contents cannot be denoted by *wh-* clauses, why are (b) and (e) grammatical? (b) and (e) are grammatical because, as it occurs in those sentences, the phrase "what John believes" is not a *wh-* clause–in fact, it is not a clause at all. Rather, it means the same as "that which John believes," and this is a singular term that refers to whatever it is (a proposition, according to Vendler) which John believes. Since knowledge cannot have propositional content, the phrase "what John believes" cannot mean the same thing in (c) and (f) as it means in (b) and (e); in (c) and (f) "what John believes" is a *wh-* clause.

But if this is right, then how can we explain away the seemingly plausible arguments given above, to the effect that what one person can believe another can know, or that the contents of our knowledge are individuated more finely than by their truth-conditions? When I ask "Do you *believe* that the ship has sailed, or do you *know* it?," the pronoun "it" does not refer anaphorically to the proposition that the ship has sailed; in fact, according to Harman 2002, this pronoun is not being used referentially at all. Rather, it is what Geach 1975 calls a "pronoun of laziness." It is to be interpreted as standing in for "that the ship has sailed." (Harman offers another example of such a pronoun in "The man who gives his paycheck to his wife is better off than the man who gives it to his mistress," where "it" does not refer back to the first man's paycheck, but instead stands in for the phrase "his paycheck.")

What about the argument that the contents of our knowledge are individuated more finely than by their truth-conditions? It's true that, if John knows that Superman is in New York and James knows that Clark Kent is in New York, then John and James need not know the same thing. But that's not because John's knowledge and James's knowledge are of propositions with different truth-conditions. Rather, it's because John knows one fact–namely, where Superman is–and James knows another fact–namely, where Clark Kent is. And recall that facts, unlike propositions, don't have truth-conditions.

3 What, Besides the Things Mentioned in 2, Are the Relata of the Knowledge Relation?

Let's use the term "knowable" as a count noun, to denote those things (be they propositions or facts) that are known. We can all, presumably, agree that for one of these knowables to be known, it must stand in the knowledge relation to a knower at a time, and perhaps in a world. So the knowledge relation involves, at least, a knowable, a knower, a time, and perhaps a world. But does it have any other relata besides these? According to *contrastivists*, it does have at least one other relatum: namely, a *contrast*. But what is a

contrast, and on what grounds do contrastivists argue that any such thing is one of the relata of the knowledge relation?

Let's begin by noticing that many, if not most, knowledge ascriptions are not of the form "S knows that p", but rather of the form "S knows *wh-* . . ." (for example, "I know when the movie starts", "she knows where the plaza is", "he knows why the Senate Finance Committee rejected the public option", "they know how to speak Portuguese"). But how should we understand knowledge *wh-* . . .? Recall from section 1 above that it is standard to understand such ascriptions as follows: S knows *wh-* = p is the true answer to the question expressed by the *wh-* clause, and S knows that p. This view is propounded by Hintikka 1975, Lewis 1982, Boer and Lycan 1986, Higginbotham 1996, and Stanley and Williamson 2001, among others. But, according to Schaffer 2007, this popular view falsely predicts that whenever two questions have the same answer, knowing the answer to one question implies knowing the answer to the other. Consider, for instance, the following two questions:

(a) Is the capital of North Carolina Raleigh or Charlotte?
(b) Is the capital of North Carolina Raleigh or Los Angeles?

It's possible for someone know the answer to easy question (b) without knowing the answer to hard question (a). But since the questions have the same answer–namely, that Raleigh is the capital of North Carolina–it follows that knowing the answer to either question cannot simply consist in knowing that Raleigh is the capital of North Carolina. More generally, Schaffer claims, knowing *wh-* cannot consist in knowing that p, where p is the true answer to the question expressed by the *wh-* clause.

Schaffer 2005 and Schaffer 2007 argue that the popular attempt to reduce knowing *wh-* to *knowing that* gets things backwards. The reduction goes in the other direction: To know that p is to know the answer to some question that has p as its true answer. To know the answer to some question that has p as its answer is to know that, among the members of the set of propositions that is the semantic value of that question, the one true member is p. For instance, to know that Abraham Lincoln was born in 1809 is to know the answer to some question that has as its true answer the proposition that Abraham Lincoln was born in 1809. What question could this be? Well, it could be the question "When was Abraham Lincoln born?"; or it could be the question "Which American president was born in 1809?"; or it could be the question "What is the significance of the year 1809 in Abraham Lincoln's life?"; or it could be any number of other questions. To know that Abraham Lincoln was born in 1809 is to stand in a particular relation–the *knowledge* relation–to some such question, as well as to the proposition Abraham Lincoln was born in 1809.

Defenders of noncontrastivist orthodoxy may challenge Schaffer's argument for contrastivism by claiming that what makes question (a) above harder than question (b) is simply that, when a listener hears question (b), she gets more useful information than she does when she hears question (a). Upon hearing question (b), the listener might reason as follows: "I am being asked whether the capital of North Carolina is Raleigh or Los Angeles, so it is probably one or the other. But it is certainly not Los Angeles, so it is probably Raleigh." The same reasoning is not available to someone who hears question (a). And so, the objection goes, the reason that someone could know the answer to question (b) without knowing the answer to question (a) is not that knowing the answer to one question differs from knowing the answer to the other

but rather because hearing question (b) puts one in a better epistemic state than does hearing question (a).

In response to this challenge and several others, Schaffer and Knobe forthcoming have attempted a defense of contrastivism by appeal to a much wider data set, gathered from surveys concerning hypothetical cases. It remains to be seen whether contrastivism provides the best explanation of the large body of survey data that they have gathered.

4 How to Explain the Apparent Context Sensitivity of Knowledge Ascriptions?

What is meant by the "apparent context sensitivity of knowledge ascriptions". We can explain this by focussing on a famous passage from DeRose 1992:

> Consider the following cases:
>
> *Bank Case A*. My wife and I are driving home on a Friday afternoon. We plan to stop at the bank on the way home to deposit our paychecks. But as we drive past the bank, we notice that the lines inside are very long, as they often are on Friday afternoons. Although we generally like to deposit our paychecks as soon as possible, it is not especially important in this case that they be deposited right away, so I suggest that we drive straight home and deposit our paychecks on Saturday morning. My wife says, 'Maybe the bank won't be open tomorrow. Lots of banks are closed on Saturdays.' I reply, 'No, I know it'll be open. I was just there two weeks ago on Saturday. It's open until noon.'
>
> *Bank Case B*. My wife and I drive past the bank on a Friday afternoon, as in Case A, and notice the long lines. I again suggest that we deposit our paychecks on Saturday morning, explaining that I was at the bank on Saturday morning only two weeks ago and discovered that it was open until noon. But in this case, we have just written a very large and very important check. If our paychecks are not deposited into our checking account before Monday morning, the important check we wrote will bounce, leaving us in a *very* bad situation. And, of course, the bank is not open on Sunday. My wife reminds me of these facts. She then says, 'Banks do change their hours. Do you know the bank will be open tomorrow?' Remaining as confident as I was before that the bank will be open then, still, I reply, 'Well, no. I'd better go in and make sure.'
>
> Assume that in both cases the bank *will* be open on Saturday and there is nothing unusual about either case that has not been included in my description of it. It seems to me that (1) when I claim to know that the bank will be open on Saturday in case A, I am saying something true. But it also seems that (2) I am saying something true in Case B when I concede that I *don't* know that the bank will be open on Saturday. Yet I seem to be in no better position to know in Case A than in Case B. It is quite natural to say that (3) If I know that the bank will be open on Saturday in Case A, then I also know that it will be in Case B. (DeRose 1992, 913)

Now, in question 4, the phrase "the apparent context sensitivity of knowledge ascriptions" is intended to denote a particular phenomenon, and Bank Cases A and B are intended to serve as examples of that phenomenon. What's puzzling about this

phenomenon is brought out by the apparent conflict between the claims that DeRose calls (1), (2) and (3) in the last quoted paragraph. If we accept the conditional (3)–namely, that if I know that the bank will be open on Saturday in Case A, then I also know that it will be in Case B–then, how can my ascription of this knowledge to myself in Case A be true, while my denial of knowledge to myself in Case B is also true? To explain the apparent context sensitivity of knowledge ascriptions, in the sense in which IV requests such an explanation, is to answer this question.

Many philosophers have, in effect, rejected the question by denying (3). For instance, Fantl and McGrath 2002, Hawthorne 2004, and Stanley 2005 all argue that whether one knows that p can depend upon the (actual or perceived) costs of falsely believing that p. When the costs are higher, the other necessary conditions of knowledge are more stringent. If this view is correct, then, given that the cost of the bank's being closed on Saturday is higher in case B than in case A, it is open to us to claim that, despite the believer's being just as confident of the bank's being open in case B as in case A, it is nonetheless true that the believer knows in case A but doesn't know in case B. That is one prominent basis on which (3) has been denied.

Another basis on which (3) has been denied is given by Bach 2005, who finds implausible the stipulation that the agent is just as confident of the bank's being open in case B as she is in case A. Given the dialogue described in case B, wouldn't the agent's confidence in the bank's being open naturally tend to drop? And if his confidence didn't drop, then wouldn't this indicate that the agent's confidence in the bank's being open is not entirely rational, and therefore not knowledge-constituting? Bach, thinking that any failure on the agent's part to lower his confidence in the bank's being open would be irrational, and thereby undermine the agent's knowledge, rejects (3), and so rejects the demand to explain what we are here calling the apparent context sensitivity of knowledge ascriptions.

But while the philosophers mentioned above have rejected (3), most philosophers would not be willing to do so. They are then confronted with the following (nonexclusive) choices: they could either reject (1), or they could reject (2), or they could reject the conjunction of (1) and (2) without rejecting either conjunct (more on this third option in a moment), or, finally, they could accept both (1) and (2) and claim that the truth of both (1) and (2) is somehow, despite appearances, consistent with the truth of (3). Only sceptics adopt the first of these three options (see, for example, Unger 1975): there is nothing special about (1) that renders it less worthy of acceptance than most of our everyday knowledge ascriptions. The only basis for rejecting (1), then, would equally well serve as a basis for rejecting most of our everyday knowledge ascriptions, and to do that is nothing short of scepticism.

But what of the other three options? Let me consider them in reverse order, starting with the fourth. Philosophers who support the fourth option call themselves "contextualists," and we will follow their usage here. (I should note, however, that the term "contextualism" has been used by other philosophers to denote views that are completely different from those being discussed here, and are rather designed to respond to the Agrippan regress, see, for example, Annis 1978 and Williams 1991.) Contextualists claim that (1) and (2) are consistent: they can both be true. They can both be true because the token sentence that I utter in Case A–"I know that it'll be open"–is not, according to contextualists, the negation of the token sentence that I utter in Case B–"Well, no [namely, I don't know that the bank will be open]." The latter might seem

to be the negation of the former, but this appearance is misleading: in fact, the two can have the same truth value, say contextualists.

How can the two sentences have the same truth value? Contextualists can hold different views on this issue. We can frame the options available to a contextualist by appeal to the following prominent account of what it is for a sentence, uttered in a particular context, to be true: "If c is a context, then an occurrence of F in c is true iff the content expressed by F in this context when evaluated with respect to the circumstance of the context" (Kaplan 1989).

According to most contextualists (e.g., Cohen 1988, DeRose 1992, Lewis 1996, Heller 1999, Rieber 1998, Neta 2002, Schaffer 2005) the reason why the token sentence that I utter in Case A is not the negation of the token sentence that I utter in Case B is that the content expressed by the sentence depends upon the context in which the sentence is tokened. Knowledge-ascribing and knowledge-denying sentences do not express the same content in different contexts, these contextualists say. Precisely how the content of the sentence depends upon the context – what features of the context it depends upon, just how it depends upon them – is in dispute among these contextualists. For instance, Cohen 1988 takes the content of knowledge-ascribing and knowledge-denying sentences to depend upon a standard of justification that is relevant in that context of utterance; DeRose 1995 takes it to depend upon a standard of closeness of possible worlds that is relevant in that context; and so on. But all of these contextualists agree that knowledge-ascribing and knowledge-denying sentences express different contents in different contexts of utterance. In contrast to this popular form of contextualism, MacFarlane 2009 sketches what he calls a "nonindexical contextualism," according to which the content of knowledge-ascribing and knowledge-denying sentences does not depend upon their context of utterance, but their truth-value does so vary, because different contexts of utterance establish a different circumstance of evaluation for the expressed content.

That is an extremely brief sketch of some of the many varieties of contextualism. What unites all of them is that, when confronted with DeRose's Bank Cases, they will all claim *both* that I am saying something true in Case B when I concede that I *don't* know that the bank will be open on Saturday, and that I am saying something true in Case A when I claim to know that the bank will be open on Saturday, even though they also grant that if I know in Case A, then I know in Case B.

Now, let's consider the alternative, noncontextualist responses to the Bank Cases. In particular, we can sketch those views according to which I speak falsely (and yet appropriately) when I say, in Case B, that I do not know that the bank is open on Saturday. According to all of these responses, when I say, in Case B, that I do not know that the bank will be open tomorrow, even though what I say is false, it is nonetheless appropriate for me to say it. That's because, by saying it, I *give it to be understood* that I cannot simply rely on the bank's being open tomorrow – and what I give to be understood is both true and conversationally relevant. Had I instead said that I do know that the bank is open tomorrow, I would have given it to be understood that I can simply rely on the bank's being open tomorrow – and what I would then have given to be understood is false. This is why, according to all of the noncontextualist responses that fall into this group, I speak falsely and yet appropriately when I say, in case B, that I do not know that the bank will be open on Saturday. Where these noncontextualist views differ is in their account of the means by which I give something true to be understood in case B.

According to Stainton 2010 and Harman 2007, when I say, in Case B, that I do not know that the bank will be open tomorrow, then even though the sentence that I assertorically utter is false, it is nonetheless appropriate for me to assert it, and that is because the content of my assertion is not the same as the content of the sentence that I utter. The content of my assertion – roughly, what it is that I claim to be true – *is* clearly true, even though the sentence that I assertorically utter is not true. This can be the case since, as noted above, the content of my assertion is not always the same as, and not exclusively determined by, the content of the sentence that I utter in performing that act.

According to Rysiew 2007 and Brown 2005, when I say, in Case B, that I do not know that the bank will be open tomorrow, even though what I say is false, it is nonetheless appropriate for me to say it. That's because, by saying it, I *conversationally implicate* that I cannot simply rely on the bank's being open tomorrow – and what I give to be understood is both true and conversationally relevant. (To say that I conversationally implicate a certain content is to say that my assertion is conversationally cooperative only if I aim to give it to be understood that that content is true. See Grice 1975.) Had I instead said (truthfully) that I do know that the bank is open tomorrow, I would have conversationally implicated that I can simply rely on the bank's being open tomorrow – and what I would then have implicated is false. This, according to Rysiew and Brown, is why I speak falsely, and yet still appropriately, in Case B.

Finally, according to a position suggested by Turri 2010, when I say, in Case B, that I do not know that the bank will be open tomorrow, I speak appropriately not because the content of my assertion or of my implicature is true. Rather, I speak appropriately because, given the high stakes I face in case B, a serious and sincere utterance of "I know that the bank will be open tomorrow" in that case would not count as a *mere assertion* that I know that the bank will be open tomorrow but as an illocutionary act of a stronger kind–call it a "guarantee" that I know that the bank will be open tomorrow. But, while I might know that the bank will be open tomorrow, and while I might even be in a position to assert this, I am certainly in no position to guarantee this.

Contextualists' chief worry about all of these noncontextualist attempts to explain why I speak falsely, and yet nonetheless appropriately, in case B when I say that I don't know whether the bank will be open on Saturday is just this: While all of these noncontextualist views make it clear why it would be *in*appropriate for me to claim to have knowledge in case B, none of them posit a well-precedented mechanism to explain why it would be appropriate for me to claim *not* to have knowledge in case B.

There is one more alternative response to the bank cases that has been much discussed lately, and this is the form of relativism espoused by Richard 2004 and MacFarlane 2005. According to their view, Kaplan's account (stated above) of what it is for a sentence, uttered in a particular context, to be true is too simple. Sentence truth is not determined simply by the content of a sentence, along with its context of utterance (and the circumstance of evaluation determined by that context of utterance). Rather it is also relative to a context of assessment. The very same content expressed in the very same context of utterance (fixing one and the same circumstance of evaluation) can be true relative to one context of assessment and false relative to another. Relativists claim that, relative to any particular context of assessment, either my knowledge-ascription in case A will be false and my knowledge-denial in case B will be true, or vice versa. But, they will say, there is no context of assessment relative to which both my knowledge-ascription in case A and my knowledge-denial in case B will count as true.

References

Annis, D. (1978) "A Contextualist Theory of Epistemic Justification," *American Philosophical Quarterly* 15: 213–219.

Bach, K. (2005) "The Emperor's New 'Knows' " in G. Preyer and G. Peter (eds.) *Contextualism in Philosophy*, Oxford: Oxford University Press.

Boer, S. and Lycan, W. (1986) *Knowing Who*, Cambridge, MA: MIT Press.

Brown, J. (2005) "Comparing Contextualism and Invariantism on the Correctness of Contextualist Intuitions," *Grazer Philosophische Studien* 69: 71–100.

Cohen, S. (1988) "How to be a Fallibilist," *Philosophical Perspectives* 2: 91–123.

DeRose, K. (1992) "Contextualism and Knowledge Attributions," *Philosophy and Phenomenological Research* 52: 913–29.

—— (1995) "Solving the Skeptical Problem," *Philosophical Review* 104: 1–52.

Devitt, M. (1996) *Coming to Our Senses*, Cambridge: Cambridge University Press.

Fantl, J. and McGrath, M. (2002) "Evidence, Pragmatics, and Justification," *Philosophical Review* 111: 67–94.

Geach, P. (1975) "Back-Reference," *Philosophia* 5: 192–206.

Ginet, C. (1975) *Knowledge, Perception, and Memory*, Ithaca: Cornell University Press.

Grice, P. (1975) "Logic and Conversation" in P. Cole and J. Morgan (eds.) *Syntax and Semantics 3: Speech Acts*, New York: Academic Press.

Groenendijk, J. and Stokhof, M. (1984) Studies in the Semantics of Questions and the Pragmatics of Answers, University of Amsterdam dissertation.

Hamblin, C. (1958) "Questions," *Australasian Journal of Philosophy* 36: 159–68.

Harman, G. (2002) "Reflections on Knowledge and its Limits," *Philosophical Review* 111: 417–28.

—— (2007) "Epistemic Contextualism as a Theory of Primary Speaker Meaning," *Philosophy and Phenomenological Research* 75: 173–9.

Hawthorne, J. (2004) *Knowledge and Lotteries*, Oxford: Oxford University Press.

Heller, M. (1999) "The Proper Role for Contextualism in an Anti-Luck Epistemology," *Philosophical Perspectives* 13: 115–29.

Higginbotham, J. (1996) "The Semantics of Questions" in S. Lappin (ed.) *The Handbook of Contemporary Semantic Theory*, Oxford: Oxford University Press.

Hintikka, J. (1975) *The Intensions of Intentionality and Other New Models for Modality*, Dordrecht: Kluwer.

Kaplan, D. (1989) "Demonstratives" in J. Almog, J. Perry, and H. Wettstein (eds.) *Themes from Kaplan*, New York: Oxford University Press.

Lewis, D. (1982) "Whether Report" in T. Pauli (ed.) *Philosophical Essays Dedicated to Lennart Aqvist on His Fiftieth Birthday*, Uppsala: University of Uppsala Press.

—— (1990) "What Experience Teaches" in W. Lycan (ed.) *Mind and Cognition: A Reader*, Blackwell: Oxford.

—— (1996) "Elusive Knowledge," *Australasian Journal of Philosophy* 74: 549–67.

MacFarlane, J. (2005) "The Assessment Sensitivity of Knowledge Attributions," in T. Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology*, Oxford: Oxford University Press.

—— (2009) "Nonindexical Contextualism," *Synthese* 166: 231–50.

Neta, R. (2002) "S Knows that P," *Noûs* 36: 663–89.

Putnam, H. (1996) "Introduction" in A. Pessin and S. Goldberg (eds.) *The Twin Earth Chronicles*, Armonk, NY: M.E. Sharpe.

Richard, M. (2004) "Contextualism and Relativism," *Philosophical Studies* 119: 215–42.

Rieber, S. (1998) "Skepticism and Contrastive Explanation," *Noûs* 32: 189–204.

Russell, B. (1918) "The Philosophy of Logical Atomism," *Monist*

Ryle, G. (1949) *The Concept of Mind*, London: Methuen.

Rysiew, P. (2007) "Speaking of Knowing," *Noûs* 41: 627–62.

Schaffer, J. (2005) "Contrastive Knowledge," in T. Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology 1*, Oxford: Oxford University Press.

—— (2007) "Knowing the Answer," *Philosophy and Phenomenological Research* 75: 383–403.

Schaffer, J. and Knobe, J. (forthcoming) "Contrastive Knowledge Surveyed."

Stainton, R. (2010) "Contextualism in Epistemology and the Context Sensitivity of 'Knows'" in J.C. Campbell, M. O'Rourke, and H. Silverstein (eds.) *Knowledge and Skepticism*, Cambridge, MA: MIT Press.

Stanley, J. (2005) *Knowledge and Practical Interests*, Oxford: Oxford University Press.
Stanley, J. and Williamson, T. (2001) "Knowing How," *Journal of Philosophy* 98: 411–44.
Turri, J. (2010) "Epistemic Invariantism and Speech Act Contextualism," *Philosophical Review* 119:
Unger, P. (1975) *Ignorance*, Oxford: Oxford University Press.
Vendler, Z. (1972) *Res Cogitans*, Ithaca: Cornell University Press.
Williams, M. (1991) *Unnatural Doubts*, Princeton: Princeton University Press.

6.2
PHILOSOPHY OF LANGUAGE FOR METAETHICS

Mark Schroeder

Metaethics is the study of metaphysics, epistemology, the philosophy of mind, and the philosophy of language, insofar as they relate to the subject matter of moral or, more broadly, normative discourse—the subject matter of what is good, bad, right or wrong, just, reasonable, rational, what we must or ought to do, or otherwise. But out of these four 'core' areas of philosophy, it is plausibly the philosophy of language that is most central to metaethics—and not simply because 'metaethics' was for a long time construed more narrowly as a name for the study of moral language. The philosophy of language is central to metaethics because both the advantages of and the open problems facing different metaethical theories differ sharply over the answers those theories give to central questions in the philosophy of language. In fact, among the open problems over which such theories differ are included, particularly, further problems in the philosophy of language. This chapter briefly surveys a range of broad categories of views in metaethics and both catalogues some of the principal issues faced by each in the philosophy of language, as well as how those arise out of their answers to more basic questions in the philosophy of language. I make no claim to completeness, only to raising a variety of important issues.

1 Cognitivist-realist Theories

Probably the default view in metaethics is to assume that moral language does not differ in any deep or important way from nonnormative language, so that sentences like 'stealing is wrong' or 'it's good that Max is happy', like sentences like 'stealing is common' or 'it seems that Max is happy' are about some subject matter in the world—what is wrong or good. On the simplest version of such a view, 'it's good that Max is happy' expresses a single, invariant proposition across all contexts of utterance, but much the same issues face more careful views according to which all context-dependence of 'it's good that Max is happy', if there is any, is of an 'unsurprising' sort—due to facts such as that 'good', like 'tall', is a gradable adjective, or due to its tense.

One of the primary issues facing such metaethical theories in the philosophy of language concerns how we manage to talk about normative properties or facts. Conse-

quently, metaethical realists are particularly interested in, or at least ought to be particularly interested in, the theory of reference—*metasemantics*, as it is sometimes called. The issues of importance to metaethical realists in the theory of reference, however, depend on the realist's answer to the metaphysical question of just what normative sentences are about. According to the nonreductive metaethical realist, normative sentences are about a *sui generis* normative domain, separate from any natural fact investigatable by the rest of the natural or social sciences. Consequently, nonreductive metaethical realists have a large stake in the debate between causal and inferential-role theories of reference—if *sui generis* normative facts are causally inert, then a causal theory of reference will not enable us to think or talk about them. This drives some work by metaethical realists outside of the philosophy of language—to account for how nonreductive, *sui generis* normative facts could in fact play a causal role (see particularly Sturgeon [1985], [2006] and Thomson [1996])—and it motivates a general interest in noncausal theories of reference, for those who don't accept this conclusion (see particularly Wedgwood [2007]).

Reductive metaethical realists, in contrast, hold that normative properties or facts can be analyzed in terms of nonnormative or natural properties or facts—so that the subject matter of normative sentences is part of the world revealed to us by the natural and social sciences. Reductive realists' stake in the theory of reference is different: they must be able to explain why the analyses that they provide are surprising. Reductive metaethical realists therefore have a deep stake in understanding philosophical analysis in general and in the philosophy of language underlying surprising property identities, in particular. Much of metaethical inquiry has been concerned with this set of issues, and the most extensive work by metaethical realists in the philosophy of language has been devoted to it (see particularly Moore [1903], Boyd [1988], Brink [1989], Railton [1986], Smith [1994], and Jackson [1998]).

Reductive realists have been quick to point out that philosophy is full of surprising identities—including the identity of water with H_2O. But things are not quite so easy. It is familiar how the identity of water and H_2O could be surprising—because it is an empirical finding—and familiar how it could be an empirical finding—because 'water' could have referred to something else, if our environment had been different, even though people took it to have the same inferential significance. But it is far from clear that 'wrong' could easily have referred to something else, simply because our environment had been different and even though people took it to have the same inferential significance—on the contrary, if we imagine cases in which people call different things wrong merely because of a difference in their environment, it is fairly easy to pump the intuition that we simply disagree with these people about what is wrong. (See particularly Hare [1952], Dreier [1990], and Horgan and Timmons [1991], [1992].)

The apparent non-'twin-earthability' of 'wrong' raises general issues in the philosophy of language about whether surprising analyses like the claim that water is H_2O must be *aposteriori* (or might be synthetic without being *aposteriori*) and about whether making an *aposteriori* discovery always involves ruling out some genuine metaphysical possibility—as discovering that water is H_2O involves ruling out the possibility that we are on a twin Earth. This has interested metaethicists in whether other terms—for example, for functional kinds—exhibit similarly twin-Earth resistant behavior, and in examples of difficult-to-discover but non-*aposteriori* analyses outside of metaethics—for example, the mathematical definition of continuity (see particularly Copp [2000], van

Roojen [2006], Schroeder [2007b]). It also raises general questions from the philosophy of language about how people could succeed at talking about the same thing as one another, even while having very different ideas about the extension of those terms (see particularly Horgan and Timmons [2000], Dreier [1993]).

2 Contextualist Theories

As suggested earlier, context-dependence in moral sentences that is due to unsurprising features such as tense or the presence of gradable adjectives is something for which ordinary realist theories in metaethics can allow. But some theories—*interestingly* contextualist theories—hold not only that 'it is good that Max is happy' expresses different propositions in different contexts of utterance but that it does so even when controlling for such unsurprising sorts of context sensitivity. It is possible, of course, for a nonreductive metaethical realist to also endorse this kind of surprising contextualism, but in practice, surprising contextualist theories are used primarily to make it easier to locate normative facts in the natural world. On such views, an important part of why it does not seem to be possible to identify moral facts with any particular natural facts is that moral claims express different natural facts, in different contexts of utterance (see particularly Dreier [1990], Timmons [1999], and Finlay [2004], [2009]).

The traditional form taken by metaethical contextualism is speaker subjectivism. The most naive forms of speaker subjectivism identify the meanings of 'stealing is wrong' and 'I disapprove of stealing', for some attitude verb like 'disapprove', which picks out a negative desire-like attitude. More cautious forms identify not the meaning of these two expressions but their truth relative to contexts of utterance and assignments (see Ayer [1936] for an argument which motivates this move). More cautious forms yet allow that there may not be any word in ordinary English for the requisite desire-like attitude, so 'disapprove' needs to be understood as a stipulative technical term (see Urmson [1968] for an argument that motivates this move). Contemporary twists on speaker subjectivism include actually rigidified speaker subjectivism, and speaker-group or conversational subjectivism, which appeals to something more like 'we disapprove of stealing' (for both, see Dreier [1990]).

Another contemporary twist on speaker subjectivism is to hold that there is a function from the speaker's attitudes to the content of 'stealing is wrong' but that these contents are not themselves about the speaker's attitudes. This view can be thought of as identifying the contents of 'stealing is wrong' and something like 'stealing instantiates dthat[the property I disapprove of actions for instantiating]' (compare Barker [2000], Schroeder [2009]). Even more flexible views require that the context contribute some parameter, and allow that in some contexts this parameter is fixed by the speaker's attitudes or interests but also allow for 'disengaged' contexts in which the parameter can be fixed by other people's attitudes or interests, in the way that it is plausible that uses of 'local' refer by default to the neighborhood of the speaker's location but can also easily refer to the neighborhood of some other location (see Timmons [1999], Finlay [2009]).

Contextualists in metaethics face a set of problems familiar from other areas of philosophy. In particular, natural tests for context-dependence, including the behavior of cross-contextual indirect discourse reports and truth ascriptions, seem to yield the wrong results for many normative sentences. For example, consider the following bit of reasoning:

Immanuel (in one context):	Lying is always wrong.
Jeremy (in a different context):	Immanuel just said that lying is always wrong.
Jeremy:	But lying is not always wrong.
Jeremy:	So Immanuel just said something false.

On the face of it, Jeremy's reasoning in this scenario makes perfect sense. But from the contextualist's point of view, it is very puzzling. For either Jeremy is correctly reporting what Immanuel said or he is making an error. If he is correctly reporting what Immanuel said, then since Immanuel's context is different and hence he did not in fact say anything inconsistent with what Jeremy means by going on to say that lying is always wrong, it would seem to follow that Jeremy's inference is invalid. But it seems both that Jeremy can correctly report what Immanuel said in this way, and that he can legitimately go on to reason in this way—at one and the same time. Sophisticated contextualist views can account for the correctness of the indirect discourse report (in some contexts), and can account for the validity of the inference (in other contexts), but the hard thing is accounting for both at the same time (and in the same context—compare Schroeder [2009]).

Another important and related issue facing contextualists is how to explain how speakers are able to engage in moral disagreements with one another. This problem was particularly acute for speaker-subjectivists, for whom even speakers in the same conversation were in relevantly different contexts; focusing instead on the speaker's conversational partners allows people in the same conversation to disagree but still runs into challenges explaining disagreement across conversations (compare Dreier [1990]).

3 Relativism

An alternative to metaethical contextualism is genuine relativism. According to genuine relativism, though 'stealing is wrong' expresses the same proposition in different contexts of utterance (again controlling for possible effects of less controversial context-dependent features), that proposition is not true or false simpliciter but may be assessed differently as true or false from different perspectives (see MacFarlane [2005], [2007]). Like metaethical contextualism, relativism is compatible with nonreductive views in metaethics, but it is hard to see what the motivation for such a position would be; relativism is more naturally motivated on similar grounds to those for contextualism—the relativist may claim that the truth of a normative proposition relative to a given point of view is itself a straightforward natural matter of fact.

As a matter of the history of the literature, metaethicists have been slow to embrace and articulate genuinely relativist views; despite the historical importance of moral relativism, many careful discussions of relativism in metaethics interpret it as a kind of contextualism (compare Harman [1975], [1978], Wong [1984]). This is in part due to the fact that the resources for articulating relativism in clearly coherent ways are a more recent development in the philosophy of language. Given the historical importance of relativism in metaethics—dating at least to the presocratics—the philosophical issues surrounding relativism are particularly important for metaethics.

4 Noncognitivist Theories: Expressivism

Probably the most interesting issues arising for metaethics in the philosophy of language arise for metaethical theories which depart quite radically from conventional ideas about semantic content. These include all of the theories in the 'noncognitivist' tradition exemplified by Hägerström, Ogden and Richards, Schlick and Carnap, Ayer and Stevenson, Hare, Blackburn, and Gibbard. The most developed contemporary strand of this tradition is the contemporary theory of metaethical expressivism, according to which rather than understanding the meaning of normative sentences in terms of what propositions they express or under what conditions they are true, we should understand the meaning of normative sentences in terms of what it is to have the corresponding normative thoughts (see especially Blackburn [1993] and Gibbard [1990] and [2003]). From the expressivist's point of view, speaker subjectivists were on the right track to think that the right way to understand the meaning of 'stealing is wrong' is to compare it to 'I disapprove of stealing', but speaker subjectivists went too far to identify the meaning or even the truth conditions of these two sentences.

Compare, for example, what would happen if we identified the meanings of 'grass is green' and 'I believe that grass is green'. It follows from that view that anyone who believes that grass is green and says 'grass is green' speaks truly. This is analogous to the prediction of speaker subjectivism that anyone who disapproves of stealing and says 'stealing is wrong' speaks truly. Moreover, this view predicts that 'if I didn't believe that grass is green, then grass wouldn't be green' expresses a truth in the mouth of every speaker; this is analogous to the prediction of speaker subjectivism that 'if I didn't disapprove of stealing, then it wouldn't be wrong' expresses a truth in the mouth of every speaker. So expressivists reason that speaker subjectivists were making the very same mistake as this view about the meaning of 'grass is green' does. Since we all know that the right relationship between 'grass is green' and 'I believe that grass is green' does not lead to these problems, expressivists propose to postulate that relationship between 'stealing is wrong' and 'I disapprove of stealing'. We can stipulatively say that the first member of each pair *expresses* the attitude that the second member *reports* (see Schroeder [2008], Gibbard [2003]).

This view quickly lands expressivists in the midst of needing to answer very general questions in the philosophy of language about what the relationship *is* between 'grass is green' and the belief that grass is green. To answer these questions, expressivists need to take stands about the nature and norms on assertion and on the primacy of mental versus linguistic content. It also leads expressivists to a very different conception of the primary goals of semantic inquiry—whereas conventional theories take the goal of a semantic theory to be to assign truth values or at least truth conditions to sentences, expressivists are led to the idea that a semantic theory needs to work by associating each sentence, 'P', with what it is to think that P. This novel expressivist approach to the objectives and structure of semantic theorizing incurs deep commitments for expressivists, and the projects of justifying it and of carrying it out go to the heart of the semantic enterprise.

One potential source of solace for metaethical expressivists is the potential for applications of a similar semantic framework outside of metaethics. For example, conditionals, probability judgments, epistemic modals, and truth have all been motivated as calling for a distinctively expressivist treatment (see Adams [1975], Gibbard [1981], Price [1983], Yalcin [2011], Schroeder [2010]). On such a view, conditionals express,

rather than report, the conditional confidence of the speaker in the consequent, conditional on the antecedent; probability judgments and epistemic modals express, rather than report, the speaker's credences; and truth and falsity ascriptions express, rather than report, the speaker's agreement or disagreement with the object of the ascription. The better motivated such further applications for expressivism are, the more optimism it makes sense to have that something like the expressivist's approach to semantic theory can be carried out—so proponents and opponents of metaethical expressivism alike have a stake in all of these topics.

5 Noncognitivist Theories: Illocutionary Theories

Whereas expressivism seeks to account for the meaning of normative sentences by accounting for the nature of normative thought, an earlier generation of noncognitivist theories sought to account for the meaning of normative sentences by accounting for the nature of normative assertion. These broadly illocutionary theories seek to explain the meaning of a sentence, 'P', by saying what 'P' is used to do. Like expressivism, these theories depart from more conventional semantic approaches by declining to tell us what normative sentences are about, or what sort of state of the world would make them true—what is significant about the meaning of normative sentences, on such views, is that they are used to commend, to cajole, to discourage, or to incite—for example (see particularly Ayer [1936], Stevenson [1937], [1944], and Hare [1952]).

Illocutionary theories of normative language are obviously deeply involved in needing to resolve a wide range of issues from speech act theory; somewhat less obvious but equally important is the interest for speech act theories of an adequate understanding of the semantics of mood. Sentential mood is, after all, the conventionally mediated aspect of linguistic use that is most transparently connected with illocutionary force—though declaratives can be used to ask questions and interrogatives can be used to issue commands and imperatives can be used to make assertions, declarative mood is conventionally associated with assertion, interrogative mood is conventionally associated with questioning, and imperative mood is conventionally associated with commands or requests. Illocutionarily based metaethical theories take these facts as grounds for optimism that a semantic theory must ultimately be able to do the kinds of things that they ask of it (see Hare [1952], Smart [1984]).

Unfortunately, it isn't immediately obvious that a general theory of mood must be able to do the things that metaethical illocutionary theories require of it—for example, on many conventional approaches to mood, it is a feature only of whole sentences, and hence doesn't embed. So, although there are complex imperative sentences like 'don't steal' and 'steal everything you can', these are not cases in which imperatives embed in more complex constructions, but in which the imperative mood takes scope over a complex sentence—in other words, 'don't steal' has the structure, !!(~(you steal)), rather than the structure, ~(!!(you steal)). But importantly, normative predicates do embed in complex constructions. So if mood doesn't embed, mood is not a very useful comparison for illocutionary metaethical theorists.

However, there is some evidence that mood can embed in complex constructions. For example, there is some evidence that conditional interrogatives like 'if it's raining, then why am I not wet?' are not interrogatives about conditionals, and that conditional imperatives like 'if you're going to the store, then bring me some milk' do not issue commands to do conditional things (compare Barker [2004]). Moreover, though it is

difficult to form conjunctions of sentences of different moods using 'and', sentences like 'close the window, for it's freezing in here' and 'it's cold in here, but leave the window open' make perfect sense and do seem to combine conjuncts of imperative and declarative moods. Advocates of illocutionary metaethical theories are wise to be particularly interested in such cases, and in whether they put pressure on an adequate semantic theory of mood to allow for ways in which conventional markers of illocutionary force like mood can in fact compose into logically more complex constructions—as normative predicates need to be able to do (compare especially Boisvert and Ludwig [2006]).

6 General Issues Facing Noncognitivist Theories

In addition to the issues in the philosophy of language of special interest to expressivists, and those of special interest to illocutionary theories, both strategies for implementing the general noncognitivist program in metaethics face a shared set of difficult issues in the philosophy of language, centered around the famous Frege–Geach Problem.

The core of the Frege–Geach Problem is that of how to account for compositionality. The central problem is that since these theories are nonextensional, they can't rely on the feature of broadly extensional theories of meaning, that negation corresponds to complementation, conjunction to intersection, and disjunction to union. This is because lacking the state of mind expressed by 'P' is not the same as having the state of mind expressed by '~P' and having both the states of mind expressed by 'P' and by 'Q' is not the same as having the state of mind expressed by 'P&Q'. Similarly (for illocutionary theories), performing the illocutionary act associated with '~P' is clearly different from simply not performing the illocutionary act associated with 'P', and performing the illocutionary act associated with 'P or Q' is clearly different from either performing the illocutionary act associated with 'P' or performing that associated with 'Q'. So if a noncognitivist theory of meaning is to explain what it is to think that ~P on the basis of what it is to think that P, or to explain what illocutionary act '~P' is suited to perform on the basis of what illocutionary act 'P' is suited to perform, the task is far from straightforward (see especially Unwin [1999], [2001] and Schroeder [2008]; see Geach [1965] for the version of this problem which gave it its name).

Moreover, all noncognitivist metaethical theories need to confront very general issues in the philosophy of logic. Whatever meaning is associated with 'not', it had better suffice to explain why for any sentence 'P', 'P' and 'it is not the case that P' are inconsistent. But the conventional way of understanding inconsistency is just truth-conditional inconsistency. Since noncognitivist theories of meaning do not work by generating truth values or truth-conditions, they have no direct way of guaranteeing that 'P' and 'it is not the case that P' cannot both be true. Consequently, they must either take a very indirect approach of trying to earn their way to this very simple result, or they must construct an alternative entire theory of inconsistency—and relatedly, of validity and logical consequence (see particularly Blackburn [1984], Gibbard [1990], [2003], van Roojen [1996], and Schroeder [2008]).

Both varieties of noncognitivist theory are also deeply committed to understanding the nature of truth. According to early proponents of noncognitivist theories, normative sentences cannot be true or false, but contemporary proponents have sought to explain how they can be true or false, by appeal to broadly deflationary accounts of truth (see especially Blackburn [1993]). This is an important move, not only because we do in ordinary discourse attribute truth or falsity to normative claims but in order to

make good on the logical relationships between normative sentences on anything like ordinary views about what logical relationships involve—even by a very indirect route. Noncognitivists are also committed to solving problems about truth-aptness that do not face other theories; for example, a noncognitivist who takes expressives like 'dammit, not stealing!' as a semantic model for the meaning of 'stealing is wrong' needs to explain why the latter is truth-apt even though the former is not (compare Dreier [1996]).

Noncognitivists also face special challenges in accounting for the semantics of attitude ascriptions and indirect discourse reports—sentences like 'Max thinks that stealing is wrong' and 'Max said that stealing is wrong'. Indeed, given the special constraints under which noncognitivist semantic theories operate, there are pressing questions, even if they can account for the simple compositionality of the 'truth-conditional' connectives, of whether they will be able to account for more general compositional issues—including fitting in with generalized quantifiers, generics, modals, tense, and mood. Since noncognitivist theories really propose a quite different way of thinking about the goals of a semantic theory, their ultimate debt is to be able to carry such a semantic program out across the board, which means that they essentially have commitments for understanding the semantics of every part of natural languages (see particularly Schroeder [2008]). Commitments for noncognitivism in the philosophy of language run deep.

7 Hybrid Theories

Some metaethical theories hold that it is not sufficient to understand the meaning of normative sentences to focus either on what such sentences are about or on what such sentences are used to do—for the meaning of normative sentences involves both a 'descriptive' content and an expressive or illocutionary component. Whereas expressivists focus on the nature of assertion in general or on expressives and illocutionarily oriented noncognitivists focus on imperatives and performatives, hybrid theorists typically take as their examples racial slur terms, and sometimes other 'thick' evaluatives. Racial slurs seem to many to be transparently associated both with a particular descriptive content (that of belonging to a certain race) and to be conventionally associated with expressing contempt or denigrating (no coincidence of etymology, here). It is the position of hybrid views that normative words are at least this much like racial slurs. (See particularly Copp [2001], Barker [2000], Boisvert [2008], and Ridge [2006], [2007].)

In fact, however, theories about the semantics of racial slurs vary widely, and so hybrid theorists are well advised to pay attention not only to slurs but also to pejoratives and epithets like 'jerk', 'asshole', and 'bastard', each of which also seems to many to be associated with some descriptive content as well as a derogatory function, and to paradigmatic 'thick' evaluatives like 'courageous' and 'kind'. Another good model for hybrid theorists to be interested in is the phenomenon of intonational sneer—of the kind associated with the claim not simply that someone is fat, but that he is 'FA-at' (compare Blackburn [1992]). Though intonation is a feature of language that it is easy to forget about when we spend most of our time as theorists comparing representations of sentences on the written page, it is very plausible that intonational sneer is conventionally determined—it works even with unnatural words, for example, in the claim that someone is 'SMA-art', in the same tone of voice. The diversity of features involved with racial slurs, pejoratives and other epithets, 'thick' evaluative terms, and intonational sneer gives hybrid theorists a variety of models to work with in seeking to understand normative language, but each of these topics needs to be much better

understood in its own right, before any hybrid theory is properly up and running (see especially Hay [2011]).

One of the most important things hybrid theorists need to understand about normative language, as well as about each of these examples from other parts of language, is how the illocutionary or expressive component of these terms' meaning projects through environments like negation, the antecedents of conditionals, attitude ascriptions, and indirect-discourse reports. A variety of the payoffs for hybrid theories turn out to turn on the hypothesis that the expressive component of normative language projects through negation and the antecedents of conditionals, but not through belief ascriptions—but it's far from clear that racial slurs work in this way, so much remains for hybrid theorist to sort out (see Schroeder [2009], Potts [2003], Hay [2011]).

8 General Issues for Metaethics from the Philosophy of Language

In addition to all of the specific issues catalogued above facing metaethicists in virtue of the specific metaethical theories which they adopt, there are general issues reated to language of broad interest to all views in metaethics. For example, any theory about 'good' or the nature of the property it expresses—whether realist or expressivist—needs to start from a proper understanding of the argument structure of 'good'—whether it expresses a monadic property of states of affairs, for example, or a relation between states of affairs and agents whom those states of affairs are good for, or a function from properties to properties, or something else altogether (compare Ziff [1960], Geach [1956], Thomson [2003], Kraut [2007]). Similarly, any theory about 'ought'—whether realist or expressivist—needs to start from a proper understanding of whether it is a context-independent propositional operator, a context-dependent propositional operator (and if so what arguments are provided by context), or whether it expresses a relation between agents and actions (compare Wedgwood [2007], Schroeder [2011], Cariani [2009]). And arguably substantive issues in both normative and metaethics can also turn on the argument structure of claims about reasons (see Schroeder [2007a]). Answering all of these questions requires addressing general issues from both linguistics and the philosophy of language.

One particular issue of general interest to metaethicists is the relationship between epistemic and deontic modals. Early treatments of deontic logic were inspired by analogies with epistemic and alethic modalities, and both cross-linguistic data and the fairly pervasive availability of both epistemic and deontic readings for the very same modals ('must', 'have to', 'may', 'ought', 'should') encourages the hypothesis that these terms make a uniform semantic contribution to both epistemic and deontic readings (see von Wright [1951], Kratzer [1977], [1981]). Depending on how this uniform semantic contribution is made out, it potentially constitutes a considerable challenge to metaethicists who think that normative or deontic language is special (see especially Finlay [2009]).

As I've been indicating in this chapter, the issues connected to metaethics from the philosophy of language are extremely diverse; I've had almost nothing to say about any of them here, only an extreme bird's-eye glance at the sweep of the territory. But it's safe to say that there is much work for philosophers of language to do in metaethics, and much work for metaethicists to do, in the philosophy of language.[1]

Note

1 Special thanks to Gillian Russell and Delia Graff Fara.

References

Acton, H.B. [1936]. 'The Expletive Theory of Morals.' *Analysis* 4: 42–45.
Adams, Ernest [1975]. *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Dordrecht: Reidel.
Ayer, A.J. [1936]. *Language, Truth, and Logic*. New York: Dover.
Barker, Stephen [2000]. 'Is Value Content a Component of Conventional Implicature?' *Analysis* 60(3): 268–279.
—— [2004]. *Renewing Meaning*. Oxford: Oxford University Press.
Blackburn, Simon [1984]. *Spreading the Word*. Oxford: Oxford University Press.
—— [1992]. 'Morality and Thick Concepts.' *Proceedings of the Aristotelian Society* suppl. vol. 66: 285–299.
—— [1993]. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
—— [1998]. *Ruling Passions*. Oxford: Oxford University Press.
Boisvert, Daniel [2008]. 'Expressive-Assertivism.' *Pacific Philosophical Quarterly* 89: 169–203.
Boisvert, Daniel, and Kirk Ludwig [2006]. 'Semantics for Nondeclaratives.' In B. Smith and E. Lepore, eds., *The Oxford Handbook of the Philosophy of Language*. Oxford: Oxford University Press.
Boyd, Richard [1988]. 'How to be a Moral Realist.' Printed in Sayre-McCord [1988], 181–228.
Brink, David [1989]. *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press.
Cariani, Fabrizio [2009]. 'The Semantics of 'Ought' and the Unity of Moral Discourse.' Ph.D. Dissertation, University of California, Berkeley.
Carnap, Rudolf [1935]. *Philosophy and Logical Syntax*. Bristol, UK: Thoemmes Press.
Copp, David [2000]. 'Milk, Honey, and the Good Life on Twin Earth.' *Synthese* 124(1): 113–137.
—— [2001]. 'Realist-Expressivism: A Neglected Option for Moral Realism.' *Social Philosophy and Policy* 18: 1–43.
Dreier, James [1990]. 'Internalism and Speaker Relativism.' *Ethics* 101(1): 6–25.
—— [1993]. 'The Structure of Normative Theories.' *The Monist* 76(1): 22–40.
—— [1996]. 'Expressivist Embeddings and Minimalist Truth.' *Philosophical Studies* 83(1): 29–51.
—— [2006]. 'Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution.' *Oxford Studies in Metaethics*, volume 1. Oxford: Oxford University Press, 217–233.
Finlay, Stephen [2004]. 'The Conversational Practicality of Value Judgment.' *Journal of Ethics* 8: 205–223.
—— [2009]. 'Oughts and Ends.' *Philosophical Studies* 143: 315–340.
Geach, Peter [1956]. 'Good and Evil.' *Analysis* 17: 33–42.
—— [1965]. 'Assertion.' *Philosophical Review* 74: 449–465.
Gibbard Allan [1981]. 'Two Recent Theories of Conditionals.' In William Harper, Robert Stalnaker, and Glenn Pearce, eds., *Ifs*. Dordrecht: Reidel.
—— [1990]. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
—— [2003]. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
Hägerström, Axel [1953]. *Inquiries into the Nature of Law and Morals*. Edited by Karl Olivecrona and translated by C.D. Broad. Stockholm: Almqvist and Wiksell.
Hare, R.M. [1952]. *The Language of Morals*. Oxford: Oxford University Press.
—— [1970]. 'Meaning and Speech Acts.' *The Philosophical Review* 79(1): 3–24.
Harman, Gilbert [1975]. 'Moral Relativism Defended.' Reprinted in Harman [2000], 3–19.
—— [1978]. 'Relativistic Ethics: Morality as Politics.' Reprinted in Harman [2000], 39–57.
—— [2000]. *Explaining Value and Other Essays in Moral Philosophy*. Oxford: Oxford University Press.
Hay, Ryan [2011]. 'Hybrid Expressivism and the Analogy Between Pejoratives and Moral Language.' *European Journal of Philosophy*, published online in *Early View*, February 2011
Horgan, Terence, and Mark Timmons [1991]. 'New Wave Realism Meets Moral Twin Earth.' *Journal of Philosophical Research* 16: 447–465
—— [1992]. 'Trouble for New-Wave Moral Semantics: The Open Question Argument Revived.' *Philosophical Papers* 21(3): 153–175.
—— [2000]. 'Copping Out on Moral Twin Earth.' *Synthese* 124(1): 139–152.
Jackson, Frank [1998]. *From Metaphysics to Ethics*. Oxford: Oxford University Press.
Kratzer, Angelika [1977]. 'What Must and Can Must and Can Mean.' *Linguistics and Philosophy* 1: 337–355.

—— [1981]. 'The Notional Category of Modality.' In H.J. Eikmeyer and H. Rieser, eds., *Words, Worlds, and Contexts*. New York: De Gruyter.
Kraut, Richard [2007]. *What is Good and Why: The Ethics of Well-Being*. Cambridge, MA: Harvard University Press.
MacFarlane, John [2005]. 'Making Sense of Relative Truth.' *Proceedings of the Aristotelian Society* 105: 321–339.
—— [2007]. 'Relativism and Disagreement.' *Philosophical Studies* 132: 17–31.
Moore, G.E. [1903]. *Principia Ethica*. Cambridge: Cambridge University Press.
Ogden, C.K., and I.A. Richards [1923]. *The Meaning of Meaning*. New York: Harcourt Brace.
Perry, R.B. [1926]. *General Theory of Value: Its Meaning and Basic Principles Construed in Terms of Interest*. Cambridge, MA: Harvard University Press.
Potts, Christopher [2003]. *The Logic of Conventional Implicature*. Oxford: Oxford University Press.
Price, Huw [1983]. 'Does 'Probably' Modify Sense?' *Australasian Journal of Philosophy* 61: 396–408.
Railton, Peter [1986]. 'Moral Realism.' *Philosophical Review* 95: 163–207.
Ridge, Michael [2006]. 'Ecumenical Expressivism: Finessing Frege.' *Ethics* 116(2): 302–336.
—— [2007]. 'Ecumenical Expressivism: The Best of Both Worlds.' *Oxford Studies in Metaethics* 2: 51–76.
Ross, W.D. [1939]. *Foundations of Ethics*. Oxford: Clarendon Press, chapter 2.
Sayre-McCord, Geoffrey, ed. [1988]. *Essays on Moral Realism*. Ithaca: Cornell University Press.
Schroeder, Mark [2007a]. 'Reasons and Agent-Neutrality.' *Philosophical Studies* 135(2): 279–306.
—— [2007b]. *Slaves of the Passions*. Oxford: Oxford University Press.
—— [2008]. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
—— [2009]. 'Hybrid Expressivism: Virtues and Vices.' *Ethics* 119(2): 257–309.
—— [2010]. 'How to be an Expressivist About Truth.' In Cory Wright and Nicholas Jang Peterson, eds., *New Waves in Truth*. New York: Palgrave Macmillan, 282–298.
—— [2011]. 'Ought, Agents, and Actions.' *Philosophical Review* 120(1): 1–41.
Searle, John [1962]. 'Meaning and Speech Acts.' *Philosophical Review* 71: 423–432.
Smart, J.J.C. [1984]. *Ethics, Persuasion, and Truth*. Oxford: Oxford University Press.
Smith, Michael [1994]. *The Moral Problem*. Oxford: Basil Blackwell.
—— [2003]. 'Neutral and Relative Value After Moore.' *Ethics* 113(3): 576–598.
Stevenson, C.L. [1937]. 'The Emotive Meaning of Ethical Terms.' Reprinted in Stevenson [1963], 10–31.
—— [1944]. *Ethics and Language*. Oxford: Oxford University Press.
Sturgeon, Nicholas [1985]. 'Moral Explanations.' In Copp and Zimmerman, eds., *Morality, Reason, and Truth*. Totowa: Rowman and Allanheld.
—— [2006]. 'Moral Explanations Defended.' In James Dreier, ed., *Contemporary Debates in Moral Theory*. Malden, MA: Blackwell.
Thomson, Judith Jarvis [1996]. 'Moral Objectivity.' In Gilbert Harman and Judith Jarvis Thomson, *Moral Relativism and Moral Objectivity*. Malden, MA: Blackwell.
—— [2003]. *Goodness and Advice*. Princeton: Princeton University Press.
Timmons, Mark [1999]. *Morality Without Foundations*. Oxford: Oxford University Press.
Unwin, Nicholas [1999]. 'Quasi-Realism, Negation and the Frege-Geach Problem.' *The Philosophical Quarterly* 49(196): 337–352.
—— [2001]. 'Norms and Negation: A Problem for Gibbard's Logic.' *The Philosophical Quarterly* 51(202): 60–75.
Urmson, J.O. [1968]. *The Emotive Theory of Ethics*. New York: Oxford University Press.
van Roojen, Mark [1996]. 'Expressivism and Irrationality.' *The Philosophical Review* 105(3): 311–335.
—— [2006]. "Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument." In R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, volume 1. Oxford: Oxford University Press, 161–193.
von Wright, G.H. [1951]. 'Deontic Logic.' *Mind* 60: 1–15.
Wedgwood, Ralph [2007]. *The Nature of Normativity*. Oxford: Oxford University Press.
Wong, David [1984]. *Moral Relativity*. Berkeley: University of California Press.
Yalcin, Seth [2011]. 'Nonfactualism about Epistemic Modality.' in Andy Egan and Brian Weatherson, eds., *Epistemic Modality*. Oxford: Oxford University Press.
Ziff, Paul [1960]. *Semantic Analysis*. Ithaca: Cornell University Press.

6.3
PHILOSOPHY OF LANGUAGE FOR METAPHYSICS

6.3.1
The Language of Causation

Eric Swanson

What parts of natural language count as parts of the language of causation? Here are four *ascriptions of singular causation*:

(1) The fall caused the vase to break.
(2) The fall was the cause of the vase's breaking.
(3) The fall was a cause of the vase's breaking.
(4) The fall was causally relevant to the vase's breaking.

Ascriptions of singular causation are used to make claims about particular instances of causation between particular causal relata. Section 1 explores some features of claims like (1)–(4).

There are also *ascriptions of general causation*, as in (5):

(5) Smoking causes lung cancer.

Sentences like (5) are used to represent a causal connection between smoking 'in general' and lung cancer 'in general,' as opposed to (6), which would be used to represent a causal connection between *Al's* smoking and *his* lung cancer.

(6) Al's smoking caused his lung cancer.

Section 2 discusses ascriptions of general causation.

Then there are expressions that might have some sort of covert causal component to their meaning:

(7) Al boiled the water.
(8) Betty's insulting Carl disgusted Dawn.

The nature of any such 'causal component' is not entirely clear. Some have gone so far as to say that (7) *just means* that Al caused the water to boil, and that (8) *just means* that Betty caused Carl to feel insulted and that Betty's insulting Carl caused Dawn's disgust. These claims are likely too strong, as we will see in section 3, but this does not show that there is *no* causal component to claims like (7) and (8). There also may be a covert causal component to sentences like

(9) Binge drinking kills brain cells.

The similarities between (5) and (9) suggest that (9) resembles ascriptions of general causation more closely than it resembles ascriptions of singular causation.

Finally there are passages the proper interpretation of which seems to require appropriate beliefs about what causes what. Knowing that the city council's fear of violence would likely cause them to refuse the demonstrators a permit helps an addressee interpret the 'they' in (10) as referring to the city council.

(10) The city councilmen refused the demonstrators a permit because they feared violence.

Knowing that advocating revolution would likely cause one's application for a permit to be refused helps an addressee interpret the 'they' in (11) as referring to the demonstrators.

(11) The city councilmen refused the demonstrators a permit because they advocated revolution. (Winograd 1972: 295)

Section 4 is devoted to the relationship between causation and the structure of discourse.

1 The Language of Singular Causation

It is controversial exactly what demands a successful theory of causation must meet. This is at least in part because it is not always clear to what extent a particular judgment about a causal claim should be explained in terms of facts about the language of causation and to what extent it should be explained in terms of facts about causation. Following Anscombe (1971a), Nancy Cartwright has even argued that

> There are a variety of different kinds of relations picked out by the abstract term 'causes' and a variety of different—correct—uses of the term for a variety of different purposes, with little of substantive content in common.
>
> (Cartwright 2006a: 56–57; see also Cartwright 1999a and Cartwright 2004a, and Machamer 2000a)

At this point no theory of causation has a claim to being the 'standard theory.' But it will be helpful to have a clear, influential analysis in mind as we discuss the language of singular causation, and David Lewis's theory in his paper (1973a) "Causation" is a good starting point.

There Lewis analyzes singular causation in terms of patterns of subjunctive conditionals, sometimes called *counterfactuals* (see Subjunctive Conditionals, Chapter 3.13, this volume). To a first approximation, he holds that *e depends causally* on *c* if and only if (12) and (13) are both true. (For important refinements, see Lewis: 1986e.)

(12) If *c* had occurred, *e* would have occurred.
(13) If *c* hadn't occurred, *e* wouldn't have occurred.

On Lewis's analysis, causation itself is the ancestral of this relation of causal dependence. Put differently: suppose that there is a *Lewisian causal chain* $<e_1, e_2, e_3, \ldots, e_n>$ such that for every natural number m that is less than n, e_{m+1} causally depends on e_m. Then for all $m < n$, e_m is a cause of e_n.

Lewis's analysis of causation, like many others, predicts that *c* was a cause of *e* in some very counterintuitive cases. For example, Caesar's death 'depends causally' on his birth. But in most if not all contexts, assertions of (14) and (15) sound very odd:

(14) Caesar's birth was the cause of his death.
(15) Caesar's birth caused his death.

This oddness doesn't worry Lewis. He acknowledges that we "sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others" (Lewis 1973a). This is not surprising: There are many independent reasons for thinking that definite descriptions in general—and so 'the cause' in particular—are heavily sensitive to conversational context (see Descriptions and Context-Sensitivity, Chapters, 3.7 and 1.10, this volume). Lewisians in particular should expect context sensitivity since counterfactuals themselves are heavily context sensitive (see again Subjunctive Conditionals, Chapter 3.13, this volume). And although they convey different presuppositions—(14) presupposes that Caesar's death had a cause, and (15) does not—it doesn't seem implausible that (14) and (15) have roughly the same truth conditions (see Presupposition and Implicature, Chapters 1.4 and 1.5, this volume).

Lewis elsewhere discusses the putative context sensitivity of 'knows' and 'knowledge' in detail (see Lewis 1996a and Philosophy of Language for Epistemology, Chapter 6.1, this volume; see also Lewis 1979c). In marked contrast he prescinds from the context sensitivity in 'the cause' and 'to cause':

> I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation.
>
> (558–559)

This position—with the assumption that one can dispel any misleading features of causal language by talking about 'a cause' or 'one of the causes'—is very common (see,

e.g., Hall 2004a: 228). It is reminiscent of John Stuart Mill's position on causes and conditions:

> Since . . . mankind are accustomed, with acknowledged propriety so far as the ordinances of language are concerned, to give the name of cause to almost any one of the conditions of a phenomenon, or any portion of the whole number, arbitrarily selected . . . it will probably be admitted without longer discussion, that no one of the conditions has more claim to that title than another, and that the real cause of the phenomenon is the assemblage of all its conditions.
>
> Mill 1843a: 403

Lewis seems to assume that we can easily talk about all the causes of a phenomenon, and Mill seems to make the slightly stronger assumption that we can easily talk about all the causes of a phenomenon *at once*: That's just what it is to talk about "the real cause of the phenomenon" in the way that Mill does.

In Swanson (2010c) I argue that these assumptions are wrong. Whether something can felicitously be cited as a cause of some effect can depend, in part, on what else has already been cited as a cause of that effect, and on what we might in the future want to cite as a cause of that effect. A Lewisian might apply my theory roughly as follows: Citing one representative of a given Lewisian causal chain as a cause of a certain effect generally crowds out other potential representatives of that causal chain to the effect, so that they can no longer felicitously be cited as a cause of that effect. Speakers are thus under pressure to choose a representative of a given causal chain that is likely to serve evolving conversational needs well.

For an example, consider Hartry Field's bomb case. Billy plants a bomb in a room. Suzy comes into the room, notices the bomb, and flees. Suzy later has a checkup and is found to be in perfect health. According to Field, Stephen Yablo, and many others, "the bomb is not a cause" of Suzy's health, although the presence of the bomb caused Suzy to flee, and Suzy's fleeing is a cause of Suzy's perfect health the next day (Yablo 2004a: 119). From this it follows that the causation relation is nontransitive. For if e_1 is Billy's planting the bomb, e_2 is Suzy's fleeing, and e_3 is Suzy's good health the next day, then unless the causation is nontransitive, it follows from the fact that e_1 is a cause of e_2 and the fact that e_3 is a cause of e_1 that e_1 is a cause of e_3. A genuine counterexample to the transitivity of causation would be a counterexample to Lewis's analysis among many others. But notice that Billy's planting the bomb makes Suzy's good health the next day quite surprising. Ordinarily, we would think that Suzy's good health should be credited to her fleeing, not to Billy's planting the bomb. These reasons and others make Billy's planting the bomb a relatively poor representative of the causal paths through it to Suzy's good health the next day. So the Lewisian who appeals to my theory to explain why some causal claims are infelicitous in certain contexts has a metaphysically neutral explanation of why it would be unusual to count Billy's planting the bomb as a cause of Suzy's good health. (And in fact it is possible to count the planting of the bomb as part of a causal chain leading to Suzy's good health, as long as the speaker makes it clear that the planting of the bomb does not crowd out other events on that causal chain.) So with appropriate attention to the context sensitivity that Lewis and Mill try to ignore, we can actually defuse an important objection to many theories of causation. Metaphysicians who take their theories to be constrained by our judgments about causal locutions should thus pay close attention to the respects in which causal

locutions are context sensitive. And this is true whether the context sensitivity of causal locutions manifests itself semantically or pragmatically (see Semantics and Pragmatics Chapter 1.2 this volume).

If any of the context sensitivity of singular causal talk does manifest itself in the semantics, there may be other important upshots for the metaphysics of causation. Note that Lewis assumes that singular causation is a binary relation, simply relating exactly one cause to exactly one effect. The burgeoning literature on the role of contrasts in the language of singular causation puts pressure on this popular assumption. For example, Jonathan Schaffer argues that the context sensitivity involved in causal contrasts is semantic, and as a result that "causation is a *quaternary, contrastive* relation: *c* rather than C^* causes *e* rather than E^*, where C^* and E^* are nonempty sets of contrast events" (Schaffer 2005a: 297). He takes this hypothesis to help capture at least some of the context sensitivity of claims about singular causation. For closely related discussion, see van Fraassen 1980b; Holland 1986a; Glymour 1986a; Hilton 1986a, 1990a, and 1995a, and the sources cited therein; Hitchcock 1996a; Woodward 2003b; Maslen 2004a; Northcott 2008a; and Schaffer 2010b. For other work on the context sensitivity of the language of singular causation see Collingwood 1937a; Anderson 1938a; Gasking 1955a; Hart 1985a: 35–37 (originally published in 1959); Gorovitz 1965a; White 1965a; Zwart 1967a; Wright 1971a; Mackie 1980b: 34–38, 119–120 (originally published in 1974); Unger 1977a and 1984a: 38, 58–60, 107; Lewis 1986c: 216; Horgan 1989a; Menzies 2004a, 2007a, and 2009a; Hall 2007a; Hitchcock 2007a and 2009a; and Maslen 2009a.

2 The Language of General Causation

Recall that mere correlation between smoking and lung cancer does not suffice for the truth of (5):

(5) Smoking causes lung cancer.

Neither is a causal connection between any one person's smoking and their lung cancer sufficient for (5) to be true. And neither is any particular instance of singular causation necessary for its truth. For example, the truth of (16) is compatible with the truth of (5).

(16) Al's smoking didn't cause him to have lung cancer.

On the basis of these differences between 'singular' and 'general' (or 'token' and 'type') causation, some philosophers distinguish between them on a metaphysical level. For example, some hold that "a distinguishing mark of a general causal sentence is that its causal relata are properties" (Menzies 1989a: 59). (See also Good 1961a, 1961b; Cartwright 1979a; Sober 1984a; and Eells 1991a.) Others think that the differences between attributions of singular and general causation do not mark any deep metaphysical differences. For example, when he discusses examples like these Lewis writes:

> Presumably those are quantified statements involving causation among particular events (or non-events), but it turns out not to be easy to match up the

> causal generalizations of natural language with the available quantified forms. A sentence of the form "C-events cause E-events," for instance, can mean any of
>
> (a) For some *c* in C and some *e* in E, *c* causes *e*.
> (b) For every *e* in E, there is some *c* in C such that *c* causes *e*.
> (c) For every *c* in C, there is some *e* in E such that *c* causes *e*
>
> not to mention further ambiguities. . . . These problems are not about causation, but about our idioms of quantification.
>
> Lewis 1973a: 558

But (as Lewis recognizes) none of (a)–(c) provide very good glosses of sentences like (5).

John Carroll offers a more promising hypothesis: Ascriptions of general causation are really *generic* ascriptions of singular causation (Carroll 1988a and 1991a). To get a sense of what this hypothesis means, consider two uncontroversial examples of generic sentences (for an extensive discussion of genericity, see Generics, Chapter 3.5, this volume):

(17) Lions have four legs.
(18) The dodo is extinct.

(17) is true even though some lions have fewer than four legs. And no particular lion's having four legs is necessary for its truth. Its exception-permitting character resembles that of (5). (18) is true even though it would be a category mistake to ascribe the property of being extinct to any particular dodo. (And besides, there are no longer any dodos to whom that property might be ascribed.)

Examples of generic sentences in the literature often include bare plurals (as in (17)) or 'kind-referring' definite descriptions (as in (18)), and ascriptions of general causation use these kinds of expressions less frequently. But consider

(19) Car accidents cause a wide range of personal injuries.
(20) The smoking of cigarettes causes nearly a half a million deaths each year.

Like (17) and (18), (19) and (20) admit exceptions and do not require anything of any particular car accident or smoking of cigarettes. And there are also uncontroversial examples of gerundive and nominalized generics (Krifka 1995a: 102–105):

(21) Smoking tobacco is a bad habit.
(22) The smoking of tobacco is a bad habit.

(For further discussion of such sentences, see Carlson 1977a: 300–301; Conrad 1982a; Chierchia 1982a, 1984a; and the other sources cited by Krifka et al.) (21) and (22) obviously resemble the ascriptions of general causation (23) and (24):

(23) Smoking tobacco causes lung cancer.
(24) The smoking of tobacco causes lung cancer.

These sentences resemble standard examples of generics in other ways, too. For example, adding 'usually' or 'typically' to (23) and (24) produces "at most a slight change of meaning" (Krifka 1995a: 9): (23), (24), and (25) mean roughly the same thing.

(25) Smoking tobacco typically causes lung cancer.

And "it is very difficult to transform a characterizing sentence into the progressive without its losing its generic character" (Krifka 1995a: 12). For example, the meanings of (26) and (27) are far from the most natural readings of (24) and (25):

(26) Smoking tobacco is causing lung cancer.
(27) The smoking of tobacco is causing lung cancer.

To be sure, none of these observations *answer* the historically important questions about ascriptions of general causation and about generic sentences. But it seems likely that sustained interaction would benefit those primarily interested in the former and those primarily interested in the latter.

3 The Hidden Language of Causation

George Lakoff (1965a) influentially argues that sentences like (28) and (29) have the same "deep structure":

(28) Floyd caused the glass to melt.
(29) Floyd melted the glass.

The truth of this hypothesis (or a hypothesis close to it) would be profoundly important to philosophical thought about causation. Among other things it would open up a whole new range of causal locutions against which to test our metaphysical theories of causation. But in response to Lakoff and others Jerry Fodor argues that "one can cause an event by doing something at a time which is distinct from the time of the event. But if you melt something, then you melt it when it melts." So although (30) is fine, (31) is odd (Fodor 1970a: 432–433).

(30) Floyd caused the glass to melt on Sunday by heating it on Saturday.
(31) # Floyd melted the glass on Sunday by heating it on Saturday.

And D. A. Cruse observes that (32) is fine although (33) is odd (Cruse 1972a: 522)

(32) John caused the reflection to move.
(33) # John moved the reflection.

On Lakoff's side see also McCawley 1968a and 1972a; and Lakoff 1972a; against Lakoffians see also Katz 1970a; Smith 1970a; Chomsky 1972a; Kac 1972a and 1972b; Shibatani 1972a; Wierzbicka 1975a; and Morreall 1976a.

From out of this debate emerged views that distinguish between "direct" or "manipulative" causation on the one hand, and "indirect" or "directive" causation on the other (see, for example Nedyalkov 1969a and 1973a; Shibatani 1973, 1973a, 1973b, and 1976a; McCawley 1976a; Talmy 1976a and 2000a; Comrie 1981a and 1985a; and Wierzbicka 1988a). James McCawley, for example, argues that

> . . . "lexical causatives" such as *kill* and transitive *open* refer to direct causation, whereas corresponding periphrastic causatives such as *cause* to *die* and *cause* to open are unspecified as to the kind of causation, since the [Gricean] cooperative principles would dictate the use of the lexical causative where direct causation is involved.
>
> (McCawley 1978a: 257)

(On Gricean cooperative principles see Presupposition and Implicature, Chapters 1.4 and 1.5 this volume.) Masayoshi Shibatani writes that this distinction is "Perhaps the single most important semantic distinction linguists make in accounting for different causative forms" (Shibatani 2000b: 11). And yet he observes that "Despite its great importance, the relevant notion has not been satisfactorily defined, and grammarians have been using the terms 'direct causation' and 'indirect causation' and related ones rather vaguely without a rigorous definition" (11).

Shibatani's observation notwithstanding, it is not clear that any distinction in the area could do all the work that the 'direct'/'indirect' distinction is put to. For example, John Morreall observes that

> Causing is not an action. Killing, on the other hand, is an action; for example, it can be done quickly or slowly. . . . [So] if *kill* were derived from *cause to die*, we would expect to see sentences like
>
> (34) *John slowly caused Mary to die.
>
> But the closest we can come to a *cause to die* sentence corresponding to ['John slowly killed Mary'] is [(35)].
>
> (35) John caused Mary to die slowly.
>
> But in [(35)] it is clear that it is Mary's dying that is slow, and not John's action of killing. (Morreall 1976a: 516–517)

Responding to this argument by contending that 'direct causation' *is* an action, and so can be done quickly or slowly, does not look very promising. But if periphrastic causatives are "unspecified as to the kind of causation," as McCawley suggests in the quotation above, then there should be a reading of (34) on which 'caused' denotes 'direct' causation, and so a reading on which it means the same as 'John slowly killed Mary.' There are various ways to respond to this argument. One might, for example, reject the putative distinction between 'direct' and 'indirect' causation. Or one might reject McCawley's appeal to conversation implicature, or reject this particular derivation of 'kill.' Unfortunately I cannot assess the possible responses here.

The literature has focused on lexicalizations that seem related to ascriptions of singular causation, but lexicalizations that seem related to ascriptions of general causation

are interesting as well. For example, a Lakoffian might have thought that (9) and (36) have the same deep structure:

(9) Binge drinking kills brain cells.
(36) Binge drinking causes brain cells to die.

But paralleling Morreall's observation, (37) is fine and (38) is odd.

(37) Binge drinking slowly kills brain cells.
(38) # Binge drinking slowly causes brain cells to die.

This is again hard for Lakoffians to explain, and again appealing to a distinction between direct and indirect causation does not look very promising.

I do not hope to resolve these issues here, but just to make the point that philosophers' skills would be useful in testing how robust the putative distinctions between direct and indirect causation really are (for an opening salvo, see Thomson 1987a; but see also 1990a: 207). Philosophers' skills would also be useful in drawing any important distinctions finely enough so that they have real explanatory power. Whatever the results of such efforts, a better understanding of how the subtleties of lexical causatives reflect and influence our thoughts and intuitions about causal notions would likely benefit a wide range of philosophers interested in causation and related topics. We might even get evidence relevant to the nature of syntax and semantics (Pietroski 2005a) and to cognitive science (see, for example, Wolff 2003a and 2007a; Song 2004a; and Sloman 2005a).

4 Causation and Discourse Interpretation

There is much to be learned from considering sentences in isolation. But sentences rarely actually occur in isolation; a sentence is generally part of a discourse. And a discourse is not a motley collection of unconnected sentences: it is more or less cohesive. A discourse "forms a unified whole . . . [that is] best regarded as a semantic unit: a unit not of form but of meaning" (Halliday 1976a: 1–2). The unified nature of naturally occurring discourses is an important aspect of linguistic meaning. And so for many Frege's 'context principle'—"it is only in the context of a proposition that words have any meaning" (Frege 1980: 73, originally published in 1884)—does not go far enough. Discourses have elements of meaning that do not adhere to any particular sentence (see Anaphora, Relevance Theory, Dynamic Semantics Chapters 3.6, 1.14, and 2.7 this volume).

The extent to which a discourse strikes its interpreters as unified depends to some extent on the causal connections between the parts of the discourse that interpreters are able to perceive. Indeed, causal relations play an important enough role in discourse interpretation that it is routine to categorize discourse relations in part in terms of their relationship with causality (see, e.g., Halliday 1976a; Mann 1988a; Hobbs 1990a; Sanders 1992a; Lascarides 1993a; Kehler 2002a; and Asher 2003a). Appeal to causal relations is essential for explaining how we interpret even quite prosaic discourses. Consider the following simple example:

(39) Max opened the door. The room was pitch dark.
(40) Max switched off the light. The room was pitch dark.

Alex Lascarides and Nicholas Asher observe that (39) and (40) have similar syntax but very different "natural interpretations": "The event and state in [39] temporally overlap, whereas in ([40]) they do not" (Lascarides 1993a: 437). How should we explain this difference between (39) and (40)? Lascarides and Asher postulate a discourse relation "*Result* (α, β): The event described in α caused the event or state described in β" (440) and hypothesize that we tacitly know that "the room being dark and switching off the light, if connected, are normally such that the event causes the state" (466). The 'normally' is important here because the relevant belief is defeasible. This is to respect the fact that if (40) is embedded in a larger discourse, we can find an interpretation on which the switching off of the light and the darkness of the room do overlap:

(41) Max and Nell realized that the monkeys were all finally asleep, and decided to turn off the light in the monkey pen to conserve power. Nell stumbled around the control room trying to find the switch, but failed. Max switched off the light. The room was pitch dark. But Max was used to finding his way around the control room at all hours of the night.

The total theory that Lascarides and Asher develop is complicated enough that I will not go into its details here. The important point is that our interpretation of discourse is influenced by our conceptions of the causal relations in the world. The relationship between causation and language is extensive enough that surprisingly broad ranges of natural language are, in some sense, parts of the language of causation.

5 Conclusion

Philosophers interested in causation sometimes abstract away from the *language* of causation as though such language is likelier to mislead or to confuse than to edify, and as though we can think about causation in a way that is untainted by facts about language. To be sure, causal language does not wear its proper analysis on its sleeve, and there is much more work to be done on causal language. Nevertheless, because it is unlikely that we can cleanly excise causal thought from causal language, we should pay attention to both if we are to pay attention to either (see also The Role of Experiment and The Role of Intuitions, Chapters 4.1 and 4.6, this volume). Philosophers are well-suited to push work on the language of causation forward. And at the same time, careful attention to the language of causation could benefit a wide range of philosophical projects.

Note

Thanks to Sarah Moss for helpful discussion, and thanks to an anonymous referee for helpful comments on an earlier version of this chapter.

References

Anderson, John. 1938. "The Problem of Causality." *Australasian Journal of Philosophy*, vol.16 (2): 127–142.

Anscombe, G. E. M. 1971. *Causality and Determination: An Inaugural Lecture*. Cambridge University Press., London.

Asher, Nicholas and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.

Beebee, Helen, Christopher Hitchcock, and Peter Menzies, editors. 2009. *The Oxford Handbook of Causation*. Oxford University Press, Oxford.

Carlson, Gregory N. 1977. Reference to Kinds in English. Ph.D. thesis, University of Massachusetts, Amherst. Published in 1980 by Garland Press, New York.
Carroll, John W. 1988. "General Causation." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 1: Contributed Papers: 311–317.
—— 1991. "Property-Level Causation?" *Philosophical Studies*, vol. 63: 245–270.
Cartwright, Nancy. 1979. "Causal Laws and Effective Strategies." *Noûs*, vol. 13: 419–437.
—— 1999. *The Dappled World*. Cambridge University Press., Cambridge.
—— 2004. "Causation: One Word, Many Things." *Philosophy of Science*, vol. 71 (5): 805–819.
—— 2006. "Where Is the Theory in our 'Theories' of Causality?" *Journal of Philosophy*, vol. CIII (2): 55–66.
Chierchia, Gennaro. 1982. "Nominalization and Montague Grammar: A Semantics without Types for Natural Languages." *Linguistics and Philosophy*, vol. 5: 303–354.
—— 1984. *Topics in the Syntax and Semantics of Infinitives and Gerunds*. Ph.D. thesis, University of Massachusetts, Amherst.
Chomsky, Noam. 1972. "Some Empirical Issues in the Theory of Transformational Grammar." In *Goals of Linguistic Theory*, P. S. Peters, editor, 63–130. Prentice-Hall, Englewood Cliffs.
Collingwood, R. G. 1937–1938. "On the So-Called Idea of Causation." *Proceedings and Addresses of the American Philosophical Association*, vol. 38: 85–112.
Collins, John, Ned Hall, and L. A. Paul, editors. 2004. *Causation and Counterfactuals*. MIT Press, Cambridge., MA.
Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. The University of Chicago Press, Chicago, second edn.
—— 1985. "Causative Verb Formation and other Verb-Deriving Morphology." In *Language Typology and Syntactic Description, Volume III: Grammatical Categories and the Lexicon*, Timothy Shopen, editor, 309–348. Cambridge University Press, New York.
Conard, Bent. 1982. *Referring and Non-referring Phrases: A Study in the Use of the Gerund and the Infinitive*. Publications in the Department of English, University of Copenhagen.
Cruse, D. A. 1972. "A Note on English Causatives." *Linguistic Inquiry*, vol. 3 (4): 520–528.
Eells, Ellery. 1991. *Probabilistic Causality*. Cambridge University Press, Cambridge.
Fodor, Jerry A. 1970. "Three Reasons for Not Deriving 'Kill' from 'Cause to Die'." *Linguistic Inquiry*, vol. 1 (4): 429–438.
van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford University Press, Oxford.
Frege, Gottlob. 1980. *The Foundations of Arithmetic*. Northwestern University Press, Evanston, IL, second edn. Translated by J. L. Austin.
Gasking, Douglas. 1955. "Causation and Recipes." *Mind*, vol. 64 (256): 479–487.
Glymour, Clark. 1986. "Statistics and Causal Inference: Comment: Statistics and Metaphysics." *Journal of the American Statistical Association*, vol. 81 (396): 964–966.
Good, I. J. 1961a. "A Causal Calculus (I)." *British Journal for the Philosophy of Science*, vol. 11 (44): 305–318.
—— 1961b. "A Causal Calculus (II)." *British Journal for the Philosophy of Science*, vol. 12 (45): 43–51.
Gorovitz, Samuel. 1965. "Causal Judgments and Causal Explanations." *Journal of Philosophy*, vol. 62 (23): 695–711.
Hall, Ned. 2004. "Two Concepts of Causation." In Collins et al. (2004), 225–276.
—— 2007. "Structural Equations and Causation." *Philosophical Studies*, vol. 132: 109–136.
Halliday, M. A. K. and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman, London.
Hart, H. L. A. and Tony Honoré. 1985. *Causation in the Law*. Oxford University Press, Oxford, second edn.
Hilton, Denis J. 1990. "Conversational Processes and Causal Explanation." *Psychological Bulletin*, vol. 107 (1): 65–81.
—— 1995. "Logic and Language in Causal Explanation." In *Causal Cognition: A Multidisciplinary Debate*, David Sperber, David Premack, & Ann James Premack, editors, chap. 16, 495–525. Oxford University Press, Oxford.
Hilton, Denis J. and Ben R. Slugoski. 1986. "Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model." *Psychological Review*, vol. 93 (1): 75–88.
Hitchcock, Christopher. 1996. "The Role of Contrast in Causal and Explanatory Claims." *Synthese*, vol. 107: 395–419.
—— 2007. "Three Concepts of Causation." *Philosophy Compass*, vol. 2/3: 508–516.
Hitchcock, Christopher and Joshua Knobe. 2009. "Cause and Norm." *Journal of Philosophy*, vol. 106 (11): 587–612.

Hobbs, Jerry R. 1990. *Literature and Cognition*. Center for the Study of Language and Information, Stanford.
Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*, vol. 81 (396): 945–960.
Horgan, Terence. 1989. "Mental Quausation." *Philosophical Perspectives*, vol. 3: 47–76.
Kac, Michael B. 1972a. "Action and Result: Two Aspects of Predication in English." In Kimball (1972), 117–124.
—— 1972b. "Reply to McCawley." In Kimball (1972), 151–156.
Katz, Jerrold. 1970. "Interpretive Semantics vs. Generative Semantics." *Foundations of Language*, vol. 6: 220–259.
Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford.
Kimball, J., editor. 1972. *Syntax and Semantics*, vol. 1. Academic Press, New York.
Krifka, Manfred, Francis Jeffry Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link, and Gennaro Chierchia. 1995. "Genericity: An Introduction." In *The Generic Book*, Gregory N. Carlson and Francis Jeffry Pelletier, editors, 1–124. University of Chicago Press, Chicago.
Lakoff, George. 1965. On the Nature of Syntactic Irregularity. Ph.D. thesis, Harvard University.
Lakoff, George and John Robert Ross. 1972. "A Note on Anaphoric Islands and Causatives." *Linguistic Inquiry*, vol. 3 (1): 121–125.
Lascarides, Alex and Nicholas Asher. 1993. "Temporal Interpretation, Discourse Relations and Commonsense Entailment." *Linguistics and Philosophy*, vol. 16: 437–493.
Lewis, David K. 1973. "Causation." *Journal of Philosophy*, vol. 70: 556–567.
—— 1979. "Scorekeeping in a Language Game." In *Philosophical Papers*, vol. 1, 233–249. Oxford University Press, Oxford.
—— 1986a. "Causal Explanation." In Lewis (1986c), 214–240.
—— 1986b. "Counterfactual Dependence and Time's Arrow." In Lewis (1986c), 32–66. With postscript.
—— 1986c. *Philosophical Papers*, vol. 2. Oxford University Press, Oxford.
—— 1996. "Elusive Knowledge." In *Papers in Metaphysics and Epistemology*. Cambridge University Press, Cambridge.
Machamer, Peter, Lindley Darden, and Carl Craver. 2000. "Thinking About Mechanisms." *Philosophy of Science*, vol. 67: 1–25.
Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford University Press, Oxford, second edn.
Mann, William C. and Sandra A. Thompson. 1988. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." *Text*, vol. 8 (3): 243–281.
Maslen, Cei. 2004. "Causes, Contrasts, and the Nontransitivity of Causation." In Collins et al. (2004).
Maslen, Cei, Terry Horgan, and Helen Daly. 2009. "Mental Causation." In Beebee et al. (2009), chap. 24, 523–553.
McCawley, James D. 1968. "Lexical Insertion in a Transformational Grammar Without Deep Structure." In *Papers from the Fourth Regional Meeting, Chicago Linguistic Society*, C. J. Bailey, B. J. Darden, and A. Davison, editors, 71–80.
—— 1972. "Kac and Shibatani on the Grammar of Killing." In Kimball (1972), 139–149.
—— 1976. "Remarks on What Can Cause What." In Shibatani (1976), 117–129.
—— 1978. "Conversational Implicature and the Lexicon." In *Syntax and Semantics 9: Pragmatics*, Peter Cole, editor, 245–258. Academic Press, New York.
Menzies, Peter. 1989. "A Unified Account of Causal Relata." *Australasian Journal of Philosophy*, vol. 67 (1): 59–83.
—— 2004. "Difference-Making in Context." In Collins et al. (2004), 139–180.
—— 2007. "Causation in Context." In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Huw Price and Richard Corry, editors. Oxford University Press, Oxford.
—— 2009. "Platitudes and Counterexamples." In Beebee et al. (2009), chap. 17, 341–367.
Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive*. John W. Parker, London.
Morreall, John. 1976. "The Nonsynonymy of 'Kill' and 'Cause to Die.'" *Linguistic Inquiry*, vol. 7.
Nedyalkov, V. P. and G. G Sil'nickij. 1969. "Tipologija Morfologičeskogo i Leksičeskogo Kauzativov" [Typology of Morphological and Lexical Causatives]. In *Tipologija Kauzativnyx Konstrukcyi: Morfologičeskij Kauzativ*, A. A. Xolodovič, editor, 20–50. Nauka, Leningrad.
—— 1973. "The Typology of Morphological and Lexical Causatives." In *Trends in Soviet Theoretical Linguistics*, Ferenc Kiefer, editor, 1–32. D. Reidel Publishing Co., Dordreet

Northcott, Robert. 2008. "Causation and Contrast Classes." *Philosophical Studies*, vol. 139: 111–123.
Pietroski, Paul M. 2005. *Events and Semantic Architecture*. Oxford University Press, Oxford.
Sanders, Ted J. M., Wilbert P. M. Spooren, and Leo G. M. Noordman. 1992. "Toward a Taxonomy of Coherence Relations." *Discourse Processes*, vol. 15: 1–35.
Schaffer, Jonathan. 2005. "Contrastive Causation." *Philosophical Review*, vol. 114 (3): 297–328.
—— 2010. "Causal Contextualisms." Ms., Australian National University.
Shibatani, Masayoshi. 1972. "Three Reasons for Not Deriving 'Kill' from 'Cause to Die' in Japanese." In Kimball (1972), 125–137.
—— 1973a. "Lexical Versus Periphrastic Causatives in Korean." *Journal of Linguistics*, vol. 9: 281–297.
—— 1973b. A Linguistic Study of Causative Constructions. Ph.D. thesis, University of California, Berkeley.
—— 1973c. "Semantics of Japanese Causativization." *Foundations of Language*, vol. 9: 327–373.
——, editor. 1976. *Syntax and Semantics 6: The Grammar of Causative Constructions*. Academic Press, New York.
——. 2000. "Introduction: Some Basic Issues in the Grammar of Causation." In *The Grammar of Causation and Interpersonal Manipulation*, Masayoshi Shibatani, editor, 1–22. John Benjamins Publishing Company, Amsterdam.
Sloman, Steven. 2005. *Causal Models: How People Think about the World and Its Alternatives*. Oxford University Press, Oxford.
Smith, Carlota S. 1970. "Jespersen's 'Move and Chance' Class and Causative Verbs in English." In *Linguistic and Literary Studies in Honor of Archibald A. Hill*, M. Ali Jazayery, Edgar Polomé, and Werner Winter, editors, vol. II, 101–109. Mouton, The Hague.
Sober, Elliot. 1984. "Two Concepts of Cause." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 2: Symposia and Invited Papers: 405–424.
Song, Grace and Philip Wolff. 2004. "Linking Perceptual Properties to Linguistic Expressions of Causation." In *Language, Culture, and Mind*, Michel Achard & Suzanne Kemmer, editors, chap. 17, 237–249. CSLI Publications, Stanford.
Swanson, Eric. 2010. "Lessons from the Context Sensitivity of Causal Talk." *Journal of Philosophy*, vol. 107 (5): 221–242.
Talmy, Leonard. 1976. "Semantic Causative Types." In Shiabatani (1976), 43–116.
—— 2000. *Toward a Cognitive Semantics, Volume I: Concept Structuring Systems*. MIT Press, Cambridge MA.
Thomson, Judith Jarvis. 1987. "Verbs of Action." *Synthese*, vol. 72: 103–122.
—— 1990. *The Realm of Rights*. Harvard University Press, Cambridge MA.
Under, Peter. 1977. "The Uniqueness in Causation." *American Philosophical Quarterly*, vol. 14: 177–188.
—— 1984. *Philosophical Relativity*. University of Minnesota Press, Minneapolis.
White, Morton. 1975. *Foundations of Historical Knowledge*. Harper & Row, Publishers, New York.
Wierzbicka, Anna. 1975. "Why 'Kill' Does Not Mean 'Cause to Die': The Semantics of Action Sentences." *Foundations of Language*, vol. 13: 491–528.
—— 1988. *The Semantics of Grammar*. John Benjamins, Amsterdam.
Winograd, Terry. 1972. *Understanding Natural Language*. Academic Press, New York.
Wolff, Philip. 2003. "Direct Causation in the Linguistic Coding and Individuation of Causal Events." *Cognition*, vol. 88: 1–48.
—— 2007. "Representing Causation." *Journal of Experimental Psychology*, vol. 136 (1): 82–111.
Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.
von Wright, Georg Henrik. 1971. *Explanation and Understanding*. Cornell University Press, Ithaca.
Yablo, Stephen. 2004. "Advertisement for a Sketch of an Outline of a Prototheory of Causation." In Collins et al. (2004), 119–137.
Zwart, P. J. 1967. *Causaliteit*. van Gorcum, Assen.

6.3.2
Dispositional Expressions

Alexander Bird

1 Introduction

The class of dispositional expressions is varied and its extension is debated and potentially very broad. Goodman (1954: 3) says that it "includes not only predicates ending in 'ible' and 'able' but almost every objective predicate, such as 'is red'." We may distinguish covertly dispositional expressions: "soluble" and "fragile" for example, and overtly dispositional expressions: "disposed to dissolved when immersed in water," "disposed to break when struck." What makes these *dispositional* is, roughly speaking, that they indicate that the objects of which they are predicated would behave in certain ways (their manifestations) in certain circumstances (their stimuli).

Dispositional expressions are one kind of thing. Dispositions are another. The former is a subclass of predicates. The latter is a subclass of properties. The connection between the two is not at all straightforward. For dispositional and non-dispositional expressions may co-refer ("avarice" and "the property most conspicuously displayed by Molière's Harpagon"). And while the class of dispositional predicates is neither empty nor exhaustive, it is a matter of debate whether any or all properties are truly dispositional in nature. Paradigm dispositional expressions (for example "soluble") indicate the definite occurrence of a manifestation in suitable circumstances; the corresponding properties are "sure-fire" dispositions. In addition to these are terms (such as "irascible") that indicate some degree of liability for the manifestation to occur in suitable circumstances. Correspondingly, in addition to sure-fire dispositions there is a class of disposition-like properties that includes capacities, tendencies, and propensities.

While there has been a continuous tradition of generating and refining analyses of dispositions since the 1930s, the function and nature of such analyses have not be constant. Initially philosophers focused on analyzing dispositional expressions or concepts (linguistic or mental items), but more recent philosophical concern purports to be more with the properties themselves (features of the world). At the same time the motivation for attempts at analysis has shifted. The tradition starts with Carnap's attempt to reconcile the role of dispositional terms in science with his verificationism. In the middle part of the twentieth century, interest in dispositions centered on Ryle's dispositional view of the mind and then later on the functionalist theory of mental states. More recently, fundamental natural properties have been held to be dispositional in nature and the analysis of dispositions is invoked in accounting for the laws of nature. This 80-year history of the analysis of dispositions reveals changing conceptions of the function of philosophical analysis and its relationship to the philosophy of language.

2 Carnap on Testability and Disposition Concepts

Carnap's interest in dispositional expressions stems from a dilemma presented by his commitment to a certain brand of empiricism. That empiricism gives a special place to science. Science is the paradigm of what can be known and what can be said. Indeed the logical positivists sometimes suggest that science is coextensive with the extent of possible knowledge. According to Ayer (1936), "There is no field of experience which cannot, in principle, be brought under some form of scientific law, and no type of speculative knowledge about the world which it is, in principle, beyond the power of science to give." Given this emphasis on the primacy of science, it is important for the positivists to accommodate rather than reject the statements of science—which stands in contrast to their attitude towards the statements of metaphysics. And dispositional expressions play an important role in the statements of science. Carnap's principal example is the predicate "soluble," but many central terms in science may be considered dispositional also (although particular cases may be disputed). "Fitness" in evolutionary biology concerns an organism's disposition to survive and breed (Mills and Beatty 1979); "electric charge" denotes a body's disposition to exert a force on another charged body (Broad 1933: 267); and so forth. Thus the commitment to science implies to a commitment to provide a satisfactory philosophical account of such predicates.

This empiricism also seemed to the positivists to provide a particular role for philosophical analysis. For the all-encompassing role of science also demands, it appeared to them, that philosophy should not be regarded as making contentful statements about the world. If philosophers claim to be so doing, then their efforts would fail and should be excluded from philosophy. Hence the rejection of metaphysics. The propositions of philosophy, says Ayer (1936: 76), "are not factual but linguistic in character . . . they express definitions, or the formal consequences of definitions." Such a view requires a demarcation principle to differentiate those statements that are factual (the statements of science) from those that purport to be factual but which are not (the statements of metaphysics). Furthermore, since philosophy is in the business of supplying definitions, it invites us to supply principles that constrain what counts as an acceptable definition. As is well known, both tasks are achieved by the verification principle, which Carnap (1936: 420) expresses thus, "the meaning of a sentence is in a certain sense identical with the way we determine its truth or falsehood; and a sentence has meaning only if such a determination is possible." Like Ayer, Carnap (1935) emphasizes that philosophy is primarily concerned with language:

> The function of logical analysis is to analyze all knowledge, all assertions of science and of everyday life, in order to make clear the sense of each such assertion and the connections between them. One of the principal tasks of the logical analysis of a given proposition is to find out the method of verification for that proposition.

So logical empiricism demands that the role of philosophy is to analyze the propositions of science in terms of their method of verification, and in the light of the central role they play in science, it requires that we provide such an analysis of sentences containing dispositional predicates in particular. However, if we take the method of verification to require direct observation, then a problem arises, for the satisfaction of dispositional predicates is not always directly observable. That a particular crystal is soluble is not directly observable, so long as it is not placed in water. A sample of barium sulphate and

a sample of sodium chloride may look just the same. If neither is immersed in water, then there is no apparent difference between the solubility of the former and in the insolubility of the latter. Likewise a charged particle and a neutral particle will behave the same so long as there is no electric or magnetic field. Since an important scientific difference is not always directly verifiable, we have a *prima facie* challenge to verificationism.

While the immediate context of Carnap's article 'Testability and meaning' is the analysis of a key class of scientific predicates, logical empiricists also had a broader concern with dispositions. Phenomenalism has always been attractive to empiricists as a way of maintaining the exclusive epistemological role of experience without succumbing to skepticism. But since a simple reduction to actual experiences leads to well-known problems, unless one resorts to Berkeley's theological solution, it is natural to consider reduction to *possible* experiences, as in Mill's "permanent possibilities of sensation" or later Mach's "functional relations of elements" (where elements are, more or less, the same as sensations). It is natural to interpret Mach's relations as dispositions. Mach himself intended "function" to have its mathematical sense. But, then, Schlick (1918: 212–14) complains that we would be reducing a material thing to "something quite shadowy," and, furthermore, something that is a relation between things that do not exist (the possible but non-actual sensations).

So, in order to satisfy the verificationist demand that meaningful statements should be verifiable, the testing that will verify or confirm the presence or otherwise of a disposition needs to be more than direct observation. It might be tempting to think that "solubility" and "charge" name unobservable properties of things that cause their behavior in water or in electro-magnetic fields. But this is precisely the kind of metaphysics that positivism rejects. Instead we should understand these expressions in terms of the kind of observable test that would confirm their correct application. Clearly an appropriate test of the difference between a soluble item and an insoluble item is to observe their dissolving or not dissolving *when the items are placed in water*. So, according to Carnap (1936), a natural first pass at a definition of "*x* is soluble" is "whenever *x* is placed in water, *x* dissolves."

The use of "whenever" is a mistake, for as Mellor (1974: 106) notes, this makes solubility an immutable property. It may be that in using "whenever" Carnap was influenced by a sense of the modal nature of disposition ascriptions, to which we shall come later. Be that as it may, whether an object is soluble *now* should not depend on how it is at previous or later times, for things can gain and lose dispositions. If we ignore this error, we can express Carnap's view thus:

(D) $Sx \quad \text{iff} \quad (Wx \rightarrow Dx)$

where 'S"' symbolizes '*x* is soluble'; 'W*x*' is '*x* is placed in water; 'D*x*' is '*x* dissolves. W*x* is the *test* condition (later *stimulus* condition) and D*x* is the *response* condition (also *manifestation* condition).

The objection to (D) that Carnap immediately raises concerns an item that is never in its history placed in water, for example a match which is never placed in water and then is burned up. We see that the right-hand side of (D) is satisfied trivially. Thus the match counts as soluble even though it clearly is not.

Carnap then considers the "bilateral reduction sentence":

(R) $Wx \rightarrow (Sx \leftrightarrow Dx)$.

(R) provides a test both for the presence of and for the absence of solubility, and to that degree satisfies the requirements of the verification principle concerning meaning. On the other hand, as Carnap recognizes, (R) does not say anything about the solubility or otherwise of some item that is not in water. And so (R) does not provide any way of eliminating talk of solubility.

These points against (R) were also made by Storer (1951). Storer's response was to note that a soluble item not in water is like, in other respects, items that are in water and dissolving and is unlike, in that respect, items that are insoluble. So to say that something is soluble is to say:

> *either* it is in water and dissolves; *or* it has some property B, such that B is possessed by some other item that is in water and does dissolve and B is not possessed by any item that is in water and does not dissolve.

Storer's proposal is notable in that it is the first to introduce a component that would later become to be thought of as the *causal basis* of the disposition. This is moving away from Carnap's concern with verifiability, to the extent that the possession of the causal basis is not itself something that is guaranteed to be verifiable. For example, the sort of property that Storer had in mind would be "being sugar." Being sugar is not a directly verifiable property, and might be held to be one that is verified by the identification of its characteristic dispositions. Furthermore, without restricting the quantification over properties (for example to *natural* properties) there are substitutions for B that make any object soluble (for example the property of being identical either to this stone [insoluble, not in water] or to that sugar cube [in water and dissolving], which is a property possessed by the stone).

More immediately problematic is the fact that we can imagine that an object might have a disposition yet be the only thing of its kind to have that disposition. If it does not undergo the relevant test procedure (for example being placed in water) then Storer's analysis will deny that the object has the disposition. For example, we might imagine that industrial chemists devise an entirely new material that is soluble. The basis of its solubility is a novel feature of this material, and so no other substance dissolves in virtue of possessing this property. If the chemists only ever make one sample of this material which they burn in due course, without ever placing in water, then this material will not count as soluble. On the other hand, had the chemists made a second, identical sample, which they did place in water (and so which does dissolve), the first sample would count as soluble. In short, the objection is that whether or not an object has a disposition such as solubility cannot depend on whether other similar objects exist and are subjected to the relevant test.

3 Introducing Stronger than Material Conditionals

Carnap and his contemporaries sought to provide an analysis of dispositional concepts employing only second-order, classical, extensional logic. The suggestion that the second- order quantification should be limited to natural properties is one proposal that would break away from this restriction. A more important proposal in the same direction concerns the nature of the appropriate conditional. Ryle (1949: 123) asserts that "To say that this lump of sugar is soluble is to say that it would dissolve, if submerged anywhere, at any time and in any parcel of water." Storer (1951: 134) says concerning definitions of dispositional concepts (such as colour predicates):

> The peculiarity of all such definitions is the occurrence of sentences of the type: "If so and so *were to happen*, then such and such *would be* the case." In a current phrase, all definitions of dispositional predicates involve the use of contrary to fact conditionals.

So both Ryle and Storer recognize the connection between dispositions and counterfactuals but retreat from making much of this connection when giving further detail, primarily because of empiricist concerns at the metaphysical implications of taking counterfactuals at face value. Counterfactual conditionals (and other subjunctive conditionals, which are understood to be included) have a modal component, seemingly telling us about non-actual potentialities. Ryle takes it that there can be no fact of the matter concerning non-actual potentialities. Consequently, the sentence "this lump of sugar would dissolve if placed in water" does not assert some factual truth, such as the attribution of a property to a thing. Rather, along with law statements, such assertions must be understood as inference tickets: one is entitled to infer from "this lump of sugar is in water" to "this lump of sugar is dissolving." In effect, the modal feature of dispositions is located in the inference ticket. Ryle does not tell us what features of the world entitle us to employ such an inference ticket.

Storer points out that the second half of the second disjunct in his analysis ("B is not possessed by any item that is in water and does not dissolve") is equivalent to: everything that is B is such that if it is placed in water then it dissolves. That covers the conditional component of the counterfactual conditional, but not the modal. The modal feature is in effect what Storer is trying to capture by the idea of there being a causal basis that is shared between an item not undergoing the test and other items that are being tested. Implicitly, Storer is suggesting that the causal basis is a property that would bring about the response in the object, were it to be tested—but without resorting to modal language.

In the light of the forgoing it is not surprising that philosophers should eventually conclude that dispositions could not be analyzed using the material conditional. Rather, a stronger than material conditional needs to be employed, which we may symbolize by '$\Rightarrow$' (without saying too much about its nature). We can also use 'Sx' to denote the test/stimulus condition, 'Mx' to denote the response/manifestation condition, and '$D_{(S,M)}$' to denote the disposition to yield manifestation M in response to stimulus S.

Sellars (1958) asserts that ascription of a disposition is simply to assert a relation of implication between the stimulus and manifestation. Hence:

(W) $D_{(S,M)}x \quad \text{iff} \quad Sx \Rightarrow Mx.$

The philosophical task, according to Sellars, is to explain what '$\Rightarrow$' means. Sellars himself draws on the idea of 'causal implication' (which we find also in Pap 1958), and expresses the idea found in Storer and also in Burks (1955), that we may need to appeal to the idea that when there is dispositional relationship between Sx and Mx, there is a kind or property to which x belongs and which plays some kind of causal or nomological role in bringing about Mx. These authors were concerned whether their notions of causal implication are compatible with a Humean regularity view of causation and law (Malzkorn 2001: 343).

4 The conditional analysis of dispositions

The position in the 1950s was that philosophers recognized that dispositional and counterfactual assertions are related and that both of these have connections with statements concerning laws and causes. Goodman (1954) distinguished the analysis of counterfactuals from analyzing the meaning of law statements. On the other hand, by his own admission, Goodman was unable to articulate the details of their relationship. Furthermore, he remained committed to a Humean view of laws that distinguishes them from other regularities only in virtue of our propensity to use them in inferences and predictions (cf. the Rylean inference ticket view of dispositions and laws mentioned above).

The discussion of the analysis of dispositions was given a major impetus by the development of a semantics for counterfactuals by Stalnaker (1968) and Lewis (1973), following earlier work by Kripke on semantics for modal logic. The semantics provided for counterfactuals made them philosophically respectable, while also articulating their problematic relationship with laws. Lewis also provided an account of causation in terms of counterfactuals, allowing a further dissociation of counterfactuals, laws, and causes. Thus it was possible to see that the causal conditional of Burks, Pap, and Sellars is a conflation of two related but separate notions, the counterfactual conditional and causation.

Furthermore, Lewis's account is based on an objective Humean view of laws rather than one depending on inference tickets or habits, thus providing a firm metaphysical basis for understanding dispositions from a Humean/empiricist perspective. At the same time, concern about the empiricist credentials of any analysis was waning as a result of a more general retreat from empiricism in the philosophy of science, and the resurgence of interest in metaphysics, especially modal metaphysics, thanks to the work of Kripke and others.

In the light of the above, we can replace the '$\Rightarrow$' in (W) with a pure counterfactual conditional: '$\Box\!\!\rightarrow$'. Thus (W) becomes what would become known as the (simple) conditional analysis of dispositions, the basis of much of the recent discussion of dispositional expressions:

(CA) $D_{(S,M)}x$ iff $Sx \Box\!\!\rightarrow Mx$.

By the 1960s, it was widely accepted that dispositional statements either mean the same as or at least entail counterfactual or subjunctive conditionals. Armstrong (1969: 23) tells us, as if it were not much more than a platitude, that "If we consider the attribution of (say) brittleness to a particular piece of glass then one outstanding feature of the attribution is that it licenses certain *conditionals*. If the glass remains unbroken, then conditionals will be 'counterfactual'. If the glass had been struck, it would have broken." However, until Stalnaker and Lewis, counterfactuals were themselves sufficiently mysterious (and even suspect) that authors felt obliged not to rest content with analyzing dispositions in terms of counterfactuals but were required to bypass the counterfactuals and to give an account in yet further terms that reflect wider philosophical (typically metaphysical) concerns. After Lewis, it was possible to distinguish acceptance of the (subjunctive/counterfactual) conditional account of the meaning of disposition statements from discussion of the metaphysics of dispositions. Thus (CA) was generally accepted, while debates focused on metaphysical issues that will be mentioned below.

5 Objections to the conditional analysis and further developments

The conditional analysis, however, suffers from now well-known flaws. Martin's (1994) counterexample of a *finkish* disposition exploits the fact that in a normal case of stimulating a disposition in order to bring about its manifestation, that process takes time. The object may be stimulated at time *t*, but the manifestation occurs only at $t+\delta$. Dispositions come and go. Very hot glass is not fragile. So a fragile glass might lose its fragility by being heated. Let us imagine that were the stimulus to occur at *t*, that would cause the disposition to disappear very quickly, certainly well before $t+\delta$. As a consequence, the process that would normally lead to the manifestation is interrupted, and the manifestation does not occur. So, for example, striking a fragile glass causes it to be heated very rapidly, sufficiently rapidly that the process of breaking is halted, and the glass does not break. In such a case the glass is fragile at *t*, but, since it is struck at *t* but does not subsequently break, it is *false* at *t* that were the glass struck it would break. (See Lewis (1997) for an attempt to reform (CA) to handle finks.)

Other objections focus on the fact that even if the disposition remains in place, its presence plus the stimulus may not be causally sufficient to bring about the manifestation. For environmental conditions may need to be appropriate, and the causal process may need to take place in a particular way. If such conditions are interfered with, the manifestation may not occur. Such interferers are *masks* (Johnston 1992) or *antidotes* (Bird 1998).

The 1990s and 2000s saw increased interest in the analysis of dispositions coming from a number of quarters. In the philosophy of language, Kripke (1982) considered and rejected dispositional accounts of rule-following, meaning, and understanding. But Martin and Heil (1998) argued that such a rejection depends on accepting (CA). The falsity of (CA) means that a counterfactual account of X and a dispositional account of X will differ in certain cases. Consequently problems with a counterfactual account of, for example, intentional or free action (Frankfurt 1969) may be overcome by preferring an account in terms of dispositions or related states, such as capacities (Smith 1997). In both cases the objections can be seen to be trading on finks or masks/antidotes.

As it stands, (CA) provides an analysis only of overtly dispositional locutions, such as "is disposed to dissolve when placed in water." Nonetheless, most of the discussion of the analysis of dispositional expressions concerns covertly dispositional locutions such as "soluble," proceeding on the assumption that the equivalence between the covert and overt expressions is straightforward and obvious. For example, we found that (CA) seems to be refuted because something may be fragile, yet it is false that it would break if stressed. But such a counterexample works only on the assumption just discussed, that the covertly dispositional "*x* is fragile" is equivalent to the overtly dispositional "*x* is disposed to break when stressed." However, if that assumption is mistaken then the existence of finks and antidotes cannot be taken immediately to refute (CA). It might instead refute the simplistic equation of covertly and overtly dispositional expressions (cf. Lewis 1997: 153; Choi 2003: 576–7). Indeed Choi (2008) undertakes a defence of (CA) on precisely this basis.

A problem arises when we combine (CA) with the standard Stalnaker–Lewis semantics for counterfactuals. That semantics include the centering condition: $A \wedge B$ entails $A \mathrel{\Box\!\!\rightarrow} B$. Thus any two facts are related by the subjunctive conditional '$\Box\!\!\rightarrow$'. (CA) says that any two possible or actual states that are subjunctively related are dispositionally related. Combining these tells us that any two facts are dispositionally related. Since

that is clearly false, either (CA) needs amendment or the Stalnaker–Lewis semantics does. Since there are independent reasons for wanting to adjust the latter, it might be worth considering whether further adjustments to the semantics for '$\Box\!\!\rightarrow$' are required. For example, we may give up not only centering but also weak centering: AB entails $A \rightarrow B$. If we do this, then it is no longer clear that finks and antidotes suffice to refute (CA). For then a case where the stimulus S occurs but the manifestation M does not occur is consistent with $S \Box\!\!\rightarrow M$. That might be justified if what replaces weak centering is the condition that only in *normal* worlds where $S \Box\!\!\rightarrow M$ is true, is $S \rightarrow M$ also true (for example worlds without finks or masks/antidotes; cf. Gundersen 2002, 2004).

While such moves permit us to strengthen the tie between dispositions and subjunctive/counterfactual conditions, other developments propose movement in the opposite direction: for example the proposal that dispositions should be aligned with habitual or generic propositions (Fara 2005).

6 Metaphysical Considerations Again

Carnap's original analysis was motivated by the (anti-)metaphysical considerations that underlie his verificationism. The latter is a doctrine concerning the meanings of expressions, and so the "analysis of dispositions" is conceived of as providing a reductive account of linguistic expressions or concepts. A central concern was to avoid an analysis in terms that were equally troubling to a verificationist, such as terms for unobservable properties or stronger-than-material conditionals.

The view that without suitable analysis dispositions were to be deemed not entirely respectable remained even after the verificationist impetus to Carnap's project had waned, and counterfactual/subjunctive conditionals themselves were accepted. (CA) tells us that to predicate a disposition of an object is equivalent to asserting a subjunctive conditional of it. But that tells us little about how the world and in particular the object in question must be for the dispositional or subjunctive predications—$D_{(S,M)}\mathbf{a}$ and S**a** M**a**—to be true. Ryle's inference ticket phenomenalism rejects the idea that there is any particular way the actual world is. But as metaphysics itself returned to respectability, the question became more pressing: what would make either side of (CA) true? A common answer is that the subjunctive/counterfactual conditional is made true by the existence of some *categorical* property plus the laws of nature. Thus:

> '$D_{(S,M)}\mathbf{a}$' is made true by, for some categorical property F, F**a** and its being a law of nature that $\forall x(Fx \wedge Sx \rightarrow Mx)$

A categorical property is held to be one that has no troubling dispositional or conditional character, but is necessarily always fully manifest—shape and duration are often held to be examples, albeit disputed ones. As Mellor (1974: 157) puts it (not his own view): "Dispositions are as shameful in many eyes as pregnant spinsters used to be—ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property."

In the light of the above, the following remarks are sometimes made:

(I) The dispositional–categorical distinction is primarily a conceptual or linguistic distinction, not a metaphysical one. The dispositional–categorical distinction is not a distinction between categories of properties but between

classes of expression, between those that permit a conditional analysis and those that do not (Strawson 2008; cf. Mumford 1998: 65).

(II) Dispositional expressions ('$D_{(S,M)}$' in the above) are ones that characterize properties (such as F in the above) via their typical effects in the actual world with its actual laws of nature (Armstrong 1969, 1997, Quine 1973; Mackie 1973). Or they refer to higher-order functional properties (for example the property of possessing some categorical property such as F in the above) (Prior et al. 1982; Prior 1985). In the context of the philosophy of mind, where mental states may be seen as dispositions, the former yields an identity theory, whereas the latter yields the functionalist view.

(III) Correspondingly, the failure of (CA) thanks to finks and masks/antidotes is sometimes held (a) to undermine the distinction between dispositional and categorical *expressions*, and (b) to vindicate the *metaphysical* program of rehabilitating dispositional properties as real properties as distinct from being either just as the shadows of conditionals or identical to categorical properties or higher order functional properties realized by categorical properties (Wright 1990, Martin 1994: 7; Mumford 1998: 63; Schrenk 2010: 171; cf. Mellor 1974).

Such debates reflect some unclarity about the nature of attempts to analyze dispositions and the relationship between the semantics and metaphysics of dispositions. As the idea that philosophical analysis is a matter of investigating our concepts comes under question (Williamson 2007), it might appear that the alternative in this case will hold that the philosophical activity of analyzing dispositions concerns those properties, the dispositions, rather than our concept of disposition. But to take such a view requires that there is indeed a distinct class of things, the dispositional properties. That is what (I) above denies. (I) itself is plausible to the extent that it is plausible that (CA) or something like it is true. Conversely, if it is true that we cannot find a straightforward analysis of dispositions, it become more plausible that our dispositional expressions do pick out a distinct class of properties, properties that are dispositional in nature. The position has a rough analog in the analysis of knowledge. If the simple justified true belief account of knowledge were correct, then that would indicate that "knowledge"is a term that serves simply to pick our a subclass of beliefs, those that meet certain additional conditions. In which case, it is plausible that the analysis of knowledge is just a matter of analysis a concept. On the other hand, as Williamson (1995) argues, the failure of the attempt to analyze knowledge into belief, plus other conditions, is evidence that states of knowing are not a subclass of the states of belief but constitute a distinct kind of mental state. In which case the analysis of knowledge (conceived more broadly than simply supplying necessary and sufficient conditions) is an investigation into the nature of knowledge itself.

On the other hand there is uncertainty as to whether the dispositional and disposition-like expressions themselves form a unified class. The distinction between covert and overt dispositional expressions has already been mentioned; it is not agreed that we can assimilate these to a single class. Moreover, terms for abilities, capacities, tendencies, and propensities have similarities to disposition terms, but no clear unification of all these is yet available. Putting that variety on one side, there are questions about what sort of thing such terms could refer to. The expression "dispositional property"carries with it a degree of ambiguity. Some terms in basic physics refer to properties that are

argued to be dispositional in nature or *essentially* dispositional, for example "inertial mass," "charge," "velocity," and so forth (Ellis and Lierse 1994; Bird 2007; Lange 2005). Such properties are fundamental natural properties and so are good candidates for being sparse universals. On the other hand, the paradigm disposition expressions discussed in the literature, such as "fragility," denote properties that are neither fundamental nor obviously natural (note the diversity of things that can be fragile and the manner of their being fragile, such as both an old parchment and the economy). If it is appropriate to think of these as referring to entities at all, the referents will be abundant universals. If 'T' refers to a sparse universal, it is plausible to think that the analyzing T is analyzing the thing, the universal T, and that this is a different exercise from analyzing the concept *T* (or the term 'T'). On the other hand, if 'T' denotes an abundant universal, then it is rather less clear that there is distinction between analyzing T and analyzing the concept *T*. For the existence of abundant universals seems to be little more than an ontological shadow of the possibility of predication by the concept *T*—if, that is, there are abundant universals at all. So if properties that are paradigmatic dispositions such as fragility, solubility, malleability, etc. are not natural properties, then it is difficult to see how there can be a project of analyzing dispositions that is different from analayzing the concept *disposition*.

7 Conclusion

The earliest phase of the history of analyses of dispositions conceives itself as engaged in the task of analyzing dispositional expressions or concepts, providing substitutions equivalent in meaning. It was motivated by radically empiricist (anti-)metaphysics and epistemology which give rise to the verificationist criterion of meaningfulness. Carnap's own failed attempts show that it is impossible to provide a complete analysis of dispositional expressions that complies with that criterion. The same empiricism limits Carnap to use of truth-functional connectives, so his "conditional" analysis of dispositions invokes only the material conditional, which is too weak for the purposes required of it (for example characterizing what is true of a soluble item that is not in water and distinguishing it from an insoluble item not in water). Some authors noted that a soluble item not in water will often be similar to a soluble item that is placed in water (and so is dissolving). This approach is of interest because for the first time it introduces the idea of what we would now call the *causal basis* of the disposition, and because it raises questions concerning the naturalness of properties. Metaphysical questions of a kind inimical to radical empiricism are beginning to come to the fore. The true break with Carnap's intended program comes when it is understood that the material conditional needs to replaced by something stronger and so non-truth-functional. This eventually settles on the counterfactual/subjunctive conditional, giving the standard (simple) conditional analysis of dispositions. It settles on this slowly, partly because of residual empiricist scepticism about such conditionals but also because such conditionals were poorly understood—a state of affairs remedied by the possible-worlds semantics of Stalnaker and Lewis in the late 1960s and early 1970s.

The conditional analysis, backed up by a semantics for the conditional, is only the beginning of the contemporary story of the analysis of dispositions. This debate is pursued not, as it was by Carnap, to avoid substantive metaphysics but in order to aid it, first in order better to understand the commitments of a dispositional–functional account of mind and then thanks to dispositional essentialist accounts of natural properties and

laws. Nonetheless, the business of analyzing dispositions is still seen primarily as one concerning the nature of disposition expressions and concepts. But as such an approach is questioned by the recent debate regarding the nature of philosophical analysis itself: Is analysis concerned with concepts or with the things themselves? The intertwining of the metaphysics and semantics of dispositions suggests that the plausibility of competing answers to that question may itself depend on the outcome of the actual process of analysis itself—a process that is by no means complete.

References

Armstrong, D. M. (1969) Dispositions are causes. *Analysis* 30: 23–6.

Armstrong, D. M. (1997) *A World of States of Affairs*. Cambridge: Cambridge University Press.

Ayer, A. J. (1936) *Language, Truth, and Logic*. London: Victor Gollancz. References to the Pelican edition, Harmondsworth (1971).

Bird, A. (1998) Dispositions and antidotes. *Philosophical Quarterly* 48: 227–34.

Bird, A. (2007) *Nature's Metaphysics: Laws and properties*. Oxford: Oxford University Press.

Broad, C. D. (1933) *Examination of McTaggart's Philosophy*, Volume I. Cambridge: Cambridge University Press.

Burks, A. (1955) Dispositional statements. *Philosophy of Science* 22: 175–93.

Carnap, R. (1935) *Philosophy and Logical Syntax*. London: K. Paul, Trench, Trübner.

Carnap, R. (1936) Testability and meaning (part 1). *Philosophy of Science* 3: 420–71.

Choi, S. (2003) Improving Bird's antidotes. *Australasian Journal of Philosophy* 81: 573–80.

Choi, S. (2008) Dispositional properties and counterfactual conditionals. *Mind* 117: 795–841.

Ellis, B. and C. Lierse (1994) Dispositional essentialism. *Australasian Journal of Philosophy* 72: 27–45.

Fara, M. (2005) Dispositions and habituals. *Noûs* 39: 43–82.

Frankfurt, H. G. (1969) Alternate possibilities and moral responsibility. *The Journal of Philosophy* 66: 829–39.

Goodman, N. (1954) *Fact, Fiction, and Forecast*. London: Athlone Press. Page references to 4th edition (1983).

Gundersen, L. B. (2002) In defence of the conditional account of dispositions. *Synthese* 130: 389–411.

Gundersen, L. B. (2004) Outline of a new semantics for counterfactuals. *Pacific Philosophical Quarterly* 85: 1–20.

Johnston, M. (1992) How to speak of the colors. *Philosophical Studies* 68: 221–63.

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Lange, M. (2005) How can instantaneous velocity fulfill its causal role? *Philosophical Review* 114: 433–68.

Lewis, D. K. (1973) *Counterfactuals*. Oxford: Blackwell.

Lewis, D. K. (1997) Finkish dispositions. *Philosophical Quarterly* 47: 143–58.

Mackie, J. L. (1973) *Truth, Probability, and Paradox*. Oxford: Oxford University Press.

Malzkorn, W. (2001) Defining disposition concepts: A brief history of the problem. *Studies in History and Philosophy of Science* 32: 335–53.

Martin, C. B. (1994) Dispositions and conditionals. *Philosophical Quarterly* 44: 1–8.

Martin, C. B. and J. Heil (1998) Rules and powers. *Philosophical Perspectives* 12: 283–311.

Mellor, D. H. (1974) In defense of dispositions. *Philosophical Review* 83: 157–81.

Mills, S. K. and J. H. Beatty (1979) The propensity interpretation of fitness. *Philosophy of Science* 46: 263–86.

Mumford, S. (1998) *Dispositions*. Oxford: Oxford University Press.

Pap, A. (1958) Disposition concepts and extensional logics. In H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science vol. 2*, pp. 196–224. Minneapolis, MN: University of Minnesota Press.

Prior, E. (1985) *Dispositions*. Aberdeen: Aberdeen University Press.

Prior, E., R. Pargetter, and F. Jackson (1982) Three theses about dispositions. *American Philosophical Quarterly* 19: 251–257.

Quine, W. (1973) *The Roots of Reference*. Lasalle, IL: Open Court.

Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson.

Schlick, M. (1918) *Allgemeine Erkenntnislehre*. Berlin: Julius Springer. References to the translation by A. Blumberg, *General Theory of Knowledge*, published by Open Court, La Salle, IL, (1985).

Schrenk, M. (2010) Antidotes for dispositional essentialism. In A. Marmodoro (ed.), *The Metaphysics of Powers, Their Grounding and Their Manifestations*. Abingdon: Routledge.
Sellars, W. (1958) Counterfactuals, dispositions, and the causal modalities. In H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science vol. 2*, pp. 225–308. Minneapolis, MN: University of Minnesota Press.
Smith, M. (1997) A theory of freedom and responsibility. In G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*. Oxford: Clarendon Press.
Stalnaker, R. (1968) A theory of conditionals. In *Studies in Logical Theory*. Oxford: Blackwell.
Storer, T. (1951) On defining 'soluble'. *Analysis* 11: 134–37.
Strawson, G. (2008) The identity of the categorical and the dispositional. *Analysis* 68: 271–82.
Williamson, T. (1995) Is knowing a state of mind? *Mind* 104: 533–65.
Williamson, T. (2007) *The Philosophy of Philosophy*. London: Wiley–Blackwell.
Wright, A. (1990) Dispositions, antirealism and empiricism. *Proceedings of the Aristotelian Society* 91: 39–59.

6.4
PHILOSOPHY OF LANGUAGE FOR NORMATIVE ETHICS

6.4.1
Language, Gender, and Sexuality

Sally McConnell-Ginet

1 Why Might We Care?

Interest in the interaction of language with gender and sexuality has primarily arisen from moral concerns about unjust social arrangements based on sex, sexual orientation, or gender. There are certainly people interested in questions in this area who take women and men to differ profoundly from one another and see gender arrangements in general, including their linguistic dimensions, as having nothing to do with social justice and everything to do with biologically fixed natures. Similarly, some see gays and lesbians as morally deviant and do not see arrangements that tilt towards straights as involving any kind of morally questionable bias. And even those who do think that sexism and homophobia are morally repugnant may pursue inquiry into the intersection of language, gender, and sexuality that is not focused on ethical dimensions. While not assuming that all men are better off than all women (certainly not the case) or that sexual minorities everywhere face problematic discrimination (again, matters are more complicated), this discussion will emphasize the relation of language to power and privilege, its relation to people's well-being as members of communities as well as its cognitive impact.

On the one hand, language can be seen as a mirror of society and culture. It can help us better understand gender and sexual arrangements and can provide important material for demystifying sexism, heterosexism, and related structures deeply embedded in familiar ways of saying things. On the other hand, language can also be viewed as actually doing things, as supplying tools and weapons, resources that benefit or harm. The suspicion is that women and sexual minorities suffer much of the harm and that straight men are most likely to reap benefits. The second perspective, which takes language as not simply indicative but partly constitutive of social injustice, is more controversial

but ultimately more illuminating. Controversy about the possibility of language as a tool or weapon often results from restricting *language* to linguistic structures, which in and of themselves are causally empty. Where language actually does enter into human affairs, it is because linguistic practices are fundamentally intertwined with other social practices both at local community and larger institutional levels as well as important for framing and sustaining ideologies. And, importantly, some of those connections are mediated by cognitive processes.

1.1 Semantic Derogation and Pejoration Over Time

It is fairly easy to see how linguistic phenomena can be taken as reflecting sexism or homophobia and then used in diagnosing the unjust social arrangements supporting these biases. For example, many words for women that were once evaluatively neutral have degenerated into slurs, usually with sexual overtones, whereas words for men, though they do occasionally undergo pejoration, do not do so with such frequency and are far less sexualized. *Hussy*, e.g., was once just a familiar form of *huswif* 'housewife'; it now has an archaic flavor but in my own girlhood it was an insult aimed at girls and women who were judged 'forward' or 'impudent' and who either in words or other deeds indicated a lack of respect for traditional sexual 'morality'—i.e., for women, chastity, passivity, and deference to men.

Many readers probably know that it's been centuries since *hussy* acquired its censorious meaning. Presumably, that particular word only gives evidence of past attitudes towards women and their sexuality, attitudes importantly inflected by class and other social factors. Etymology tells us nothing about current gender arrangements. It does, however, tell us something about common assumptions at a particular point in the past (early to mid-17th c) about those labeled *huswif* that must have made it possible for people to implicate conversationally by *hussy* what the word now conveys conventionally (McConnell-Ginet 1989 offers a 'just-so' account of this shift). But few contemporary speakers even use *hussy*, much less connect it to *housewife*. Is the example relevant? Although things certainly have changed, studies of terms for designating women and men conducted very recently continue to document problematic attitudes—to be labeled *stud* often counts as an achievement whereas the label *slut* is seldom embraced. And it is the wide-scale patterning of which the particular *huswif-hussy* shift is an example that is really significant, not the particular case.

1.2 Language as Tool or Weapon

Less obviously, linguistic practices can be understood as in part constitutive of unfair gender and sexual practices, sometimes directly and sometimes through helping strengthen the attitudes and ideologies that sustain them. I will briefly consider examples of each.

1.2.1 Epithets and slurs

Epithets seem to work directly, not just to express contempt and hostility but to enact them. *Bitch* is sometimes used to refer to arbitrary girls and women but it is also used to attack particular people. Uppity women who might be asserting themselves forcibly can expect to hear *you bitch*, and the association of the slur with female strength and resistance has led some women to embrace the designation for themselves. But the form

is also used to impose servility—*you're my bitch and don't forget it*—few women would want to reclaim that usage. Such an utterance can also occur as an extremely offensive move from one male prisoner to another, an assertion of dominance and sexual control and in that context often a verbal prelude to rape. This kind of usage mixes sexism and homophobia, a toxic combination that recurs with other slurs.

The once unprintable *cunt* uses a word for female genitals to attack and here too the attack can be by a man on another man. (The word can also be used just to designate women in general, a usage that is probably most common in hypermasculine contexts like locker-rooms—here the harm comes in indirectly via the way in which such practices sustain sexist ideologies.) Many find this form so thoroughly offensive that they cannot say it even to provide an example just as some find the vitriolic force of the "N-word" leaking through even in indirect speech or didactic contexts (see Anderson & Lepore, forthcoming). Though the form *prick* uses a term designating male genitals to insult, its force is relatively mild, and it is not used to refer to men in general but just those with particular kinds of characteristics. Furthermore, there is no general practice of attacking or insulting women by applying male-designating terms to them. Men are, however, insulted even by terms that in their application to women are not directly insulting: *girl*, *lady*, *woman*—and it is overwhelmingly men who issue such insults to other men or boys, usually to impugn their masculinity, which is often taken to require heterosexuality as well as aggression, toughness, and the like. The asymmetry in insult potential of terms designating females and males is related to the general phenomenon, discussed in the next section, of equating the 'normal' human condition with maleness. Epithets explicitly targeting sexual minorities are also often used to police gender and sexual conformity. *Dyke* and *faggot* express hostility when used by those who do not identify as lesbians or gay men, although, like *bitch*, in-group reclamation is possible as in Alison Bechdel's comic strip, *Dykes to Look Out For*.

Most people find it easy to see slurs hurled at someone directly as harming or at the very least upsetting that person. But it is not easy to understand just how this works. As a child, my friends and I would intone "sticks and stones may break my bones, but words will never hurt me," as tears ran down our cheeks in the face of verbal onslaughts from peers (*fatty*, *four-eyes*, and *nigger-lover* were among those I remember trying with my chant—unsuccessfully, of course—to render powerless). Expanding on (and modifying somewhat) ideas in Kaplan 1999 and the comments on them in Kratzer 1999, Potts 2007 has proposed that epithets and 'expressive' language generally involve a dimension of content orthogonal to the primary or 'at-issue' content of the utterance. Although this approach has virtues, it does not get at the 'visceral' character of at least some slurs, their resistance to rendering in other terms, their capacity to inflict psychic wounds, none the less real for not involving broken bones or the like. The idea of backgrounded negative content does, however, help explain why slurs used even in the absence of their targets might nonetheless be deeply harmful because they help 'naturalize' the highly negative attitudes that underlie them, treating them as defaults that the speaker presumes are shared by everyone not in the target group. The harm in such 'naturalizing' effects is indirect in the sense of not being immediately realized in a particular utterance situation but such subtle support of problematic ideologies, which is not only hard to confront but often passes unnoticed, is probably ultimately even more consequential than direct harm of the kind delivered by targeted slurs .

1.2.2 Masculine generics

A major target of feminist-backed language reform campaigns in the 1970s and 1980s was *he/man* language of the sort promoted in English classes for more than a century. This includes the use of the pronominal forms *he/him/his* not only in contexts where potential referents are presumed male but also in contexts where referents might be female (as in i).

i. Any student having trouble in a course should contact his instructor about whether to drop.

It also includes the use of *man/men* to designate both the species *homo sapiens* (without an article as in ii) and (with an article as in iii or without as in iv) to designate male human beings.

ii. Man is thought to be the only species that anticipates death.
iii. A strange man just knocked on the door.
iv. Man without woman will not flourish.

And there are contexts where it is unclear whether *man* designates the species generally or just the male members thereof. Sentence v can be interpreted as equivalent to iv but it need not be, having a possible more inclusive interpretation.

v. Man without a mate will not flourish.

(These examples also seem to presume heterosexuality, although it may now be possible to construe *mate* as potentially covering a 'life partner' of the same sex.) As many observed (and a few empirical studies confirmed), *he/man* talk often rendered women less visible than if more inclusive language were used—e.g., for pronouns, *he or she* or singular *they* (often frowned on in high school grammar classes but used by Shakespeare and Jane Austen along with many others) or for the species, *homo sapiens* or *humans*.

Many resisted complaints about generic masculines, explaining that it was 'just a grammatical convention'—certainly, many English users have been drilled relentlessly in the prescription to use *he* in contexts where women or girls are potential referents since grammarian Ann Fisher formulated it in 1745 as the way to avoid using *they* with a singular antecedent, a widespread usage she deplored. (*You*, which takes a plural verb form and was originally opposed to singular *thou*, did not arouse grammarians' ire when used with singular reference, perhaps in part because the shift predated the rise of prescriptive grammars aimed at foreigners and those with relatively little education, including many women.) Robin Lakoff (1973, 1975), the first linguist to argue at length that American women were linguistically disadvantaged, nonetheless thought *he* unproblematic as did novelist Ursula LeGuin in *The Left Hand of Darkness*, her fantasy novel envisioning a world where people's sex might change often. Both Lakoff and LeGuin have since shifted their initial positions as they have become persuaded that this 'convention' is not quite what it seems and is closely connected to a more general phenomenon of viewing (adult) men as canonical exemplars of humanity, women and children being marked and somewhat deviant cases.

Unfortunately, switching to more inclusive forms of language—using *they* as a pronoun for antecedents that do not specify sex, e.g.—does not mean women are automatically included. Even when apparently generic forms are used, empirical research finds a tendency to forget about potential female referents except in stereotypically female contexts (e.g., when the antecedent of *they* is *nurse*). Notice what is going on with what Treichler and Frank 1989 dub 'false generics', which were first noted though not so labeled by Black and Coward 1981.

vi. We awoke in the morning to find that the villagers had all paddled away in their canoes, leaving us behind with the women and children.
vii. The pioneers and their wives were often ill-equipped to deal with the bone-chilling winds and the heavy snows they experienced once winter began.

Words like *villager* and *pioneer* certainly do not seem to have masculinity as part of their semantic content and yet apparently writers can presume that the women and children do not 'really' count as potential referents.

Philosopher Elizabeth Beardsley (1976) argued that 'genderization' is the crux of the problem, by which she meant specifying sex when it ought to be irrelevant as in *poetess* or *woman doctor*. As Beardsley correctly noted, such specifications tend, by familiar Gricean principles, to suggest that a poet's or a doctor's being female is noteworthy (whereas being male is expected)—and sometimes as well that the poet or doctor who is female is somehow not quite 'up to' the standard set by men. Unfortunately, matters are more complicated than we might at first think, given that supposedly inclusive forms like *poet* and *doctor* also tend to suggest men unless something is done to forestall those suggestions—e.g., talking about Emily Dickinson or Virginia Apgar (the physician who devised what is now known as the Apgar Test, a quick assessment of infant health used in American delivery rooms). These complications do not mean that it is unimportant to avoid *he/man* language, but they do mean that there is no simple strategy for both including women and also avoiding irrelevant marking of gender. There have to be multiple approaches and contexts matter in decisions over which approach to take for a particular utterance.

Part of the problem in understanding such phenomena is that what is 'brought to mind linguistically' includes more than what is covered by the familiar Gricean trio of expression meaning, utterance meaning, and speaker meaning. As they process words uttered, hearers regularly make inferences and build cognitive representations, often not at readily accessible conscious levels, that contribute to later discourse in the same way as content added to the common ground on the basis of understanding meaning that has been linguistically conveyed or communicated by implicature.

2 'Hearer Meaning': Significance Beyond Meaning and Implication

On the model of expression, utterance, and speaker meaning, I will use Hearer Meaning to include other stuff that linguistic practices may 'bring to mind'. Conceptual Baggage covers inferences triggered for hearers by the situated use of particular words. Emergent Significance is more global, covering what hearers glean, often with little or no direct awareness, on the basis of significant patterns of utterances across many different discourses (or perhaps sometimes, significant patterns over the history of a single exchange). Emergent significance arises from processing what others have said—i.e.,

develops in a hearer's mind—but this kind of 'hearer's meaning' may manifest itself when that individual speaks through what s/he presupposes, takes for granted.

2.1 Conceptual Baggage

Back in the early 1970s some feminists told a story along these lines.

> A young boy was rushed to the hospital from the scene of an accident in which his father was killed and prepped for emergency surgery. The surgeon walked in, took one look, and said "I can't operate on him—that's my son."

Many people, including many self-identified feminists, were puzzled. How could the boy be the surgeon's son when his father had been killed? They were assuming that the surgeon was male and overlooking the possibility that the surgeon was the boy's mother. Though recognizing the existence of a stereotype that surgeons are male (perhaps less robust now than when the story was first told but by no means completely moribund), semanticists tended to cringe when people took the story to illustrate that *surgeon* is semantically 'male'.

I just recently noticed other assumptions, which are associated with *father*—people also don't tend to entertain the possibility that the boy had two fathers, which can happen a number of ways, including having two gay fathers. It's not the definite *his* that triggers this ignoring of the possibility of two fathers as can be seen by substituting *brother* for *father*. Just as *surgeon* triggers assumptions about the maleness of that profession, so *father* triggers assumptions about the canonical heterosexual family in which a child only has one father. (There can, of course, also be multiple fathers if a child has a stepfather or an adoptive father and also a biological father.)

What McConnell-Ginet 2008 proposed is that words have associated with them cognitively not only information about their linguistic properties, including their conventional meaning, but also further conceptual baggage. This would include stuff like the stereotypes mentioned in Putnam 1973 as well as frames and scripts (essentially, as Fillmore 1982 points out, frames that include temporal structure). Cognitive and computational linguists find such notions critical for modeling language use and understanding: WordNet, e.g., really builds on frame analyses. And in thinking about many examples important for understanding issues of the representation of gender and sexuality, we find conceptual baggage lurking in the background. Although conceptual baggage is not part of what the words conventionally mean or, usually, of what the speaker means in uttering the words, it can be essential for understanding how what is said using those words contributes to the inferences interlocutors draw. At least some times, conceptual baggage may get loaded onto a word because of the emergent significance of discursive patterns in which the word figures.

2.2 Emergent Significance

Discussions of language, gender, and sexuality often talk about the emergent significance of discursive patterning, although not in those terms. Analysts draw inferences not from a single utterance or even a single situated discourse but from widespread statistical patterns across many different discourses. For example, children's books, syntax texts, newspapers, and many other media include many more references to men and

boys than to women and girls. (See Macaulay & Brice 1994 for the syntax texts and references to other domains; Gill 2007 is an excellent general discussion of media and gender, with many useful mentions of representational studies.) Notice that no particular utterer or utterance need have 'meant' that male human beings are more important than female or even more interesting or less problematic to discuss nor does anyone have to embrace such beliefs explicitly. Indeed, many people who themselves contribute to these patterns might be dismayed to realize that they have done so. I speak from experience: the example sentences in my dissertation featured John and Bill, whom I thanked for their help in my final paragraph. Why not Joan or Mary? I suspect that, at some subliminal level, I thought they would be obtrusive, would not serve my need for 'ordinary' people, given their 'special' or 'marked' status as women. I would never have endorsed the explicit claim that only men were 'ordinary' people. Nonetheless, I fell in unthinkingly with discursive patterns giving undue representation to men, patterns that are very likely part of what raises the visibility of men as villagers, pioneers, and the like and decreases women's visibility overall.

Frequency of appearance is only part of the picture: males are much more often than females represented as agents, as doers, whereas females, when present, often appear as patients, recipients of others' actions. Even when speaking of women prominent in public arenas, there are still frequent comments on their appearance, their marital and parental status, and the like, whereas talk of men in similar circumstances far less often includes such personal details in the absence of their special relevance. (See again Macaulay & Brice 1994 as well as Gill 2007 for references.) Things have changed since the dark ages when I was writing my dissertation. Feminism and activism around issues of sexuality have had an impact: media representations these days often pay at least lip service to gender equity, though often combining calls for 'girl power' with pictures of hypersexualized 8- and 9-year-old girls or talk about women putting their own needs first with injunctions to present oneself 'attractively' (along with ads for costly 'beauty products', clothes, and so on). It is also notable that men are often represented in far from flattering ways—swilling beer, drooling over beautiful women, and generally behaving like immature jerks. There is also far more use of knowing 'winks' and ironic distancing than was once the case—'yes, yes, we've heard feminist complaints and have now moved on'. So discursive patterns these days are more mixed and less clear than they once were, but the gender ideologies on which they draw and to which they contribute are, to a considerable extent, still profoundly problematic.

Heteronormative presumptions can also be seen in contrasts between discourses. Kitzinger 2002 reports that she and her female partner were regularly asked on entering the dining room at a hotel where they were staying, "Are you expecting to be joined?" or "Are you ladies alone?" Noticing this practice and not wanting to jump to conclusions, Kitzinger stationed herself so as to be able to observe exchanges with other entering guests: mixed-sex couples, she found, were shown to their tables without inquiries about possible others. The ideological load of the entry question does not emerge if one simply considers it in isolation but only by comparing its presence with the absence of such a query to mixed-sex couples. What is standard for greeting guests at mealtimes is a welcome and questions like "would you like a window table?"—that general pattern is what makes the questions to the pair of women about potential other tablemates significant.

Not only are presumptions of heterosexuality still quite widespread and often not explicitly noticed. Views about what is standard in heterosexual interactions also often continue surprisingly skewed. Although women's sexual agency is more frequently

recognized than it once was, there is still a view of women as essentially passive in typical heterosexual encounters, men initiating and sustaining contact. At the same time discursive patterns of various kinds support the idea that this supposed passivity masks general receptivity: women, such patterns suggest, generally want men's sexual attentions but they do not (and should not) solicit them actively (though they are often encouraged to behave so as to interest men sexually but not to appear to be 'doing' anything purposive).

It is only against this kind of backdrop that one can make sense of widespread beliefs that a woman who has not vocally and persistently said "no" has 'implicitly consented' to a man's sexual advances even if one believes her claim that she was afraid of the man and did not really welcome the sexual activity he initiated. Has a man who hands his wallet to someone he believes can (and will) hurt him badly if he does not do so 'consented' to being robbed? Or has a male prison inmate who lets himself be sodomized when he thinks the alternative is being stabbed engaged in 'consensual' sexual activity? Susan Ehrlich, a Canadian discourse analyst, has published a number of detailed and insightful studies examining transcripts from university sexual assault hearings and criminal court proceedings on sexual assault or rape charges. Although neither Canada nor the US any longer requires evidence of 'utmost resistance' to sustain a charge of rape, Ehrlich 2002 shows patterns within particular hearings and across a number of different hearings that seem to assume 'implied consent' by a woman who does not keep repeating her objections to a sexual encounter.

Ehrlich 2007 goes further and examines differing judicial assessments of testimony from a woman complainant in a Canadian sexual assault case. The 1995 trial judge and the 1998 Alberta appeals court ruled in favor of the male defendant on the grounds that the female complainant failed to be explicit and forceful enough in her objections to the defendant's actions. They did this while at the same time saying that they believed her testimony that (1) she did not want the sexual activity to continue and that (2) she was fearful of what might happen to her if she continued her protests. When the case was appealed to the Supreme Court, however, the charge was upheld, with the judge writing the 1999 decision articulating a very different conception of *consent* than that shown by lower courts' decisions. An extract from the decision (Ehrlich 2007, 270):

> The question of implied consent should not have arisen. The trial judge's conclusion that the complainant implicitly consented and that the Crown failed to prove lack of consent was a fundamental error given that he found the complainant credible, and accepted her evidence that she said 'no' on three occasions and was afraid. *This error does not derive from the findings of fact but from mythical assumptions. It denies women's sexual autonomy and implies that women are in a state of constant consent to sexual activity. The majority of the Court of Appeal also relied on inappropriate myths and stereotypes.* (italics added by Ehrlich)

The myths and stereotypes at issue are, I suspect, still circulated, not so much by explicit representations of women as 'always wanting it' but by the emergent significance of discursive patterns implicitly presenting women's sexual passivity as a norm and their reluctance to admit to their own desires a common phenomenon.

Roger Dworkin (1966) quotes various 'experts' in arguing that 'consent' is virtually impossible to establish in cases of allegations of sexual assault or rape (and that some standard of 'resistance' is needed to establish lack of consent). He cites approvingly (p. 682) the following claim: "Although a woman may desire sexual intercourse it is *customary* for her to say 'no, no, no' (although meaning 'yes, yes, yes') and to expect the man

to be the aggressor." (italics added) Of course occasionally a woman might say 'no' and mean 'yes', but viewing this as 'customary' even in 1966 (when it could well have happened somewhat more often than it now would) showed more attention to myth than to realities. Dworkin even speaks of 'unconscious consent', a nearly incoherent notion

Now Dworkin was writing over four decades ago. Few law review authors would now put matters as he did. The Canadian trial and appeals court decisions, however, are much more recent. And they are not isolated examples of views of women's sexual passivity and 'coyness' as normative but fit into wider patterns, which we will see in the next section can constrain women's effectiveness as speakers.

3 Speech Act Effectiveness

Effectiveness is a matter of how a particular speech act is received, what changes it brings about in the ongoing course of affairs, including the conversational record of what interactants are mutually assuming as well as attitudes and actions that might result. If I issue a directive to you, do you undertake to comply, take on the indicated obligation? If I make an assertion, do you add its content to your belief set or at least strengthen your confidence in the proposition I convey? If I make a joke, do you laugh? If I ask a question, do you take on the job of trying to answer it? If I make a suggestion, do you consider it seriously? And so on.

In trying to "do things with words," some are more consistently successful than others. Success depends on many factors but the relation of the interlocutors, including their views of one another, is extremely important. The response or uptake to an utterance is critical; immediate and long-term effects on others must be considered. Grice 1957 initially defined what is was for speaker S to mean that p to hearer H in terms of S's intending to produce in H the belief that p—i.e., to produce a certain kind of response. Similarly, Austin in his 1960 account of illocutionary acts called for a certain kind of 'uptake' on the hearer's part: in order for an assertion to come off successfully, the hearer need not come to believe what the speaker asserts but must at a minimum recognize that the speaker is trying to bring it about that the hearer so believe.

Both Grice and Austin made success in meaning or in performing illocutionary acts depend on others than the speaker. Many critics noted that one could not 'intend' to do what one did not believe one could do and yet I might well mean that p to you even if I have no expectation that you could be led to believe p. So far as it goes, this criticism has some merit, but it does not really come to grips with the fact that semantic 'success' is not a purely individualistic matter: others play an important role. Although both the Gricean and the Austinian account do have problems, they are not just confused as some have assumed. A speaker whose attempts to mean are impeded by others—e.g., by others who simply refuse to believe that she means what she says when she rejects a man's sexual advances—is semantically handicapped, illocutionarily constrained. I made this point in McConnell-Ginet 1989, building on a related observation made in an unpublished talk I'd heard philosopher Sara Ketchum make in the early 1980s.

Philosophers Jennifer Hornsby and Rae Langton make a similar point when they argue (in, e.g., Hornsby & Langton 1998 and elsewhere) that the notion of 'speech' that is involved in advocacy of 'freedom of speech' is illocutionary. They suggest that speakers' freedom of speech is curtailed if they cannot get their communicative intentions recognized, cannot fully perform certain illocutionary acts because they cannot get hearers to recognize certain illocutionary intents. This point is developed in the context

of arguing that pornography tends to produce a communicative climate in which women's attempts to refuse men's sexual advances are simply not recognized as such because it is assumed women's putative refusals are just a sham and that the women saying 'no' are really communicating 'yes but you should keep on insisting'. I disagree with Hornsby and Langton that pornography is the main purveyor of this message. As I suggested in the preceding section, this view of 'normative' heterosexual encounters is implicit across a range of discourses, including legal ones. The explicitness of pornography may, in my view, render it less powerful in fostering and sustaining such pernicious views than more respectable kinds of discourse in which these ideas are only implicit (at least in recent years).

But the fundamental point is independent of whether pornography really does have the effects Hornsby and Langton posit. Communicative acts are not purely individualistic—at a minimum their success requires recognition by the hearer. And what seems clear is that recognition may be impeded by gender and sexual ideologies that the hearer draws on in assessing the communicative intentions of particular speakers. Without appropriate uptake from the hearer—recognition of the speaker's illocutionary intent—the speech act is doomed.

Of course, effectiveness requires more. When I make a sincere assertion I want you to believe what I have said, a directive is issued with an interest in the hearer's doing what it enjoins. Formal theories of discourse usually incorporate these desired outcomes in their models of the impact of utterances on developing conversation: e.g., an assertion adds the propositional content asserted to the conversational record, a directive places the requested hearer action on the hearer's "to-do" list, etc. (See, e.g., Roberts 2004 for discussion.) In other words, the hearer's Updating the conversational record or context as intended is fundamental to the speaker's illocutionary goals. Dynamic semantics does not assign propositional content to declarative sentences but rather their "context-change potential." Assuming a possible worlds account, a declarative sentence has the potential to narrow down the set of worlds still considered 'live options' (thinking propositionally, the assertion can add to the stock of propositions taken for granted) in the discourse context. Whether this potential is actually realized when a sentence is uttered depends on social factors operative in the particular discourse, among which the hearer's relation to the speaker, including his views of her interests and abilities, may often play a major role.

It is sometimes said that the desired outcomes involving, e.g., hearers' beliefs or intentions to act are not part of what a speaker means or of the illocutionary force of her utterance but are rather further perlocutionary effects of the same kind as impressing or frightening or surprising. But the kinds of outcomes we are discussing are far more intimately tied to meaning as is evident from their role in dynamic semantics: speakers aim (or at least hope) to update the discursive context in certain standard ways by their utterances. Their capacity to effect such updating, however, is socially constrained.

So both uptake—recognition of what the speaker means—and appropriate updating—making the indicated changes in the discourse context—are essential for what I will call Full Semantic Effectiveness. It is not only gender and sexuality that can limit semantic effectiveness, of course, but it is in the context of thinking about issues like women's difficulties in refusing some men's sexual advances (and also in some other communicative attempts) that I have seen the inadequacy of treating meaning without reference to hearers' uptake and updating.

As noted above utterance effectiveness is not limited to semantic effectiveness: the hearer might offer the required uptake (comprehension) and make the appropriate contextual updates yet further (perlocutionary) effects desirable to the speaker might not ensue. A speaker offers a suggestion in a business meeting and it is indeed adopted but the higher-ups adopting it attribute it to a male colleague: the speaker is not credited. Apportioning credit (or blame) for discursive moves requires keeping accurate tabs on who has said what and folks are not always good at that, especially when what was said (including the uptake and update) might not have been expected from that particular speaker. And of course further aims can be frustrated in many other ways as achieving them usually involves speaker's and hearer's assessments of one another's immediate situations and longer-term interests and goals as well as of factors independent of discourse participants. But the first essential step is semantic effectiveness, both uptake and updating. Securing the intended uptake is crucial for illocutionary agency—one must be understood—but if uptake goes well but the intended updating does not occur, the illocutionary act is seriously deficient.

4 Conclusion

I have largely ignored discussion of the relation of linguistic variation (different ways of saying 'the same thing') to gender and sexual identities. And I have also not considered such matters as conversational dynamics, amount of talk in public contexts, and other staples of the literature on language, gender, and sexuality. My emphasis has been on the explicit content of what people say and write, hear and read, and the implicit content speakers communicate as well as that generated by hearers. I have also pointed to the profoundly social character of "doing things with words" that becomes apparent when looking at the interaction of language with gender and sexuality.

Note: Section 3 draws heavily from "Linguistics and gender studies," my contribution to the forthcoming *Handbook of the Philosophy of Linguistics*, ed. R. Kempson, N. Asher, and T. Fernando in the Philosophy of Science series published by Elsevier under the general editorship of D. Gabbay, P. Thangard, and J. Woods.

References

Anderson, Luvell and Ernest Lepore. Forthcoming. Slurring words. *Noûs*.

Beardsley, Elizabeth Lane. 1976 Referential Genderisation. In Carol C. Gould and Max W. Wartofsky, eds., *Women and Philosophy* (N.Y.: G.P. Putnam's Sons), 285–93.

Black, Maria and Rosalind Coward. 1981. Linguistic, social and sexual relations. *Screen Education* 39:111–133. Rpt. in abridged form in D. Cameron, ed., *The feminist critique of language: A reader* (Routledge, 1998).

Ehrlich, Susan. 2001. *Representing rape: Language and sexual consent*. London and New York: Routledge.

Ehrlich, Susan. 2007. Legal discourse and the cultural intelligibility of gendered meanings. *Journal of Sociolinguistics* 11.4: 452–477.

Fillmore, Charles J. 1982. Frame semantics. In In-Seok Yang (ed.), *Linguistics in the morning calm* (Seoul: Hanshin for Linguistic Society of Korea), 111–137.

Frank, Francine Wattman and Treichler, Paula A. 1989. *Language, gender, and professional writing: Theoretical approaches and guidelines for nonsexist usage*. New York: Modern Language Association.

Kaplan, David. 1999. The meaning of *ouch* and *oops*: Explorations in the theory of meaning as use. Unpublished UCLA ms, paper given at Cornell Conference on Theories of Context Dependency.

Kitzinger, Celia. 2002. Doing feminist conversation analysis. In Paul McIlvenny, ed., *Talking gender and sexuality* (Amsterdam/Philadelphia: John Benjamins, 2002), 49–78.

Kratzer, Angelika. 1999. Beyond *ouch* and *oops*: How descriptive and expressive meaning interact. URL

http://semanticsarchive.net/Archive/WEwNGUyO/, A comment on David Kaplan's paper, Cornell Conference on Theories of Context Dependency.
Lakoff, Robin. 1975. *Language and woman's place*. NY: Harper & Row.
Lakoff, Robin Tolmach, ed. Mary Bucholtz. 2004. *Language and woman's place: Text and commentaries*. Revised and expanded edition of Lakoff 1975, with Lakoff's original text and commentaries by her and others. New York: Oxford University Press.
LeGuin, Ursula K. 1989. *Dancing at the edge of the world: Thoughts on words, women, places*. New York: Grove Press.
McConnell-Ginet, Sally. 1979. Prototypes, pronouns and persons. In M. Mathiot, ed., *Ethnolinguistics: Boas, Sapir and Whorf revisited* (Mouton), 63–83. Rpt. w/ updated notes and references in McConnell-Ginet, 2011.
McConnell-Ginet, Sally. 1989. The sexual reproduction of meaning: A discourse-based theory. In Frank and Treichler 1989, 35–50. Rpt. in McConnell-Ginet 2011.
McConnell-Ginet, S. 2008 Words in the world: How and why meanings can matter. *Language* 84.3, 497–527
McConnell-Ginet, S. 2011. *Gender, sexuality, and meaning: Linguistic practice and politics*. Oxford: Oxford University Press.
Treichler, Paula and Francine Wattman Frank. 1989. Guidelines for nonsexist usage. In Frank & Treichler, eds., 137–278.
Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2):165–197.

6.4.2
Language and Race

Rae Langton, Sally Haslanger, and Luvell Anderson

Introduction

What is the point of language? If we begin with that abstract question, we may be tempted towards a high-minded answer: "People say things to get other people to come to know things that they didn't know before" (Stalnaker, 2002, 703). The point is truth, knowledge, communication. If we begin with a concrete question, "What has language to do with race?" we find a different point: to attack, spread hatred, create racial hierarchy. The mere practice of racial categorization is controversial: are race terms natural or social kind terms? What does categorization do to the categorized? (Outlaw, 1990; Omi and Winant, 1994; Appiah, 1996; Andreason, 1998; Kitcher, 1999; Zack, 2002; Haslanger, 2008; Glasgow, 2009). But there is worse than mere categorization to contend with.

An Alabama court says slaves are not really people: as "rational beings, they are capable of committing crimes; and in reference to acts which are crimes, are regarded as persons," but nonetheless "because they are slaves, they are incapable of performing civil acts, and, in reference to all such, they are things, not persons" (Caterall, 1926, 247). A restaurant proprietor places a sign, "Whites Only," using language to discriminate, and help enact segregation. A Nazi publisher says:

> If we do not oppose the Jews with the entire energy of our people, we are lost. But if we can use the full force of our soul that has been released by the National Socialist revolution, we need not fear the future. The devilish hatred of the Jews plunged the world into war, need and misery. Our holy hate will bring us victory and save all of mankind.
>
> (Streicher, 1943)

White fans of Babe Ruth write hate mail to rising competitor Hank Aaron, using language to assault, insult, persecute, and threaten (Kennedy, 2003, p. 20):

> Dear Nigger Henry, You are [not] going to break this record [for home runs] established by the great Babe Ruth if you can help it . . . Whites are far more superior than jungle bunnies . . . My gun is watching your every black move.

Language can be used to express racism, to attack people, to help constitute and cause racial oppression.

There may be implications for law and policy: think of the striking down of Alabama slave codes; the outlawing of the "Whites Only" sign; the execution of the Nazi propagandist; the Congressional condemnation of the "racial slurs and insults" directed at Aaron (Rosengren, 2008, 119). There may be implications for epistemology: Think of the racist perception of accent and idiolect, contributing to an unjust distribution of credibility (Fricker, 2007); the harnessing of Biblical language to envision escape from "captivity", to a "promised land" where "justice rolls down like waters and righteousness like a mighty stream" (King, 1963). Our topic is large, and fiercely discussed by political philosophers, critical race theorists, and philosophers of law. But we shall confine our focus to three issues in philosophy of language: *slurs*, or epithets; *speech acts* of racial oppression; and racial *generics*.

1 Slurs

Slurs or epithets, like those in the letters to Aaron, are expressions that function to demean or derogate people in virtue of their group membership. They derogate classes, in contrast to more individualized insults like "moron" or "idiot." Slurs generally have neutral counterparts; what can be said with a slur can often be said another way without offense. An adequate theory of slurs should account for the variation in offensive potency among slurs: "nigger" is more offensive than "cracker"; "kite" is more offensive than "ofay"; "spic" is more offensive than "honkey". There can be variation among slurs for the same group: "nigger" is more offensive than "darkie" or "coon." Besides derogatory uses, slurs are sometimes appropriated by target members in ways that mitigate or mollify a slur's offensiveness—for example the use of "nigga" by some African Americans (Kennedy, 1999/2000, 2002).

Slurs provoke important questions in the philosophy of language, though only relatively recently have they begun to receive the attention they deserve (Dummett, 1973; Kaplan, 1999; Hornsby, 2001; Williamson, 2007; Hom, 2008; Richard, 2008; Anderson and Lepore, forthcoming). Does the derogatory aspect of slurring statements affect their truth conditions? Where is the offensiveness of slurs located? Do slurs possess expressive content as well as descriptive content? How do slurs and the semantic/pragmatic distinction bear on each other? Though we cannot do justice to all of these questions, we shall take a look at some semantic and pragmatic approaches to the truth-value of slurring sentences, with their corresponding accounts of the location of a slur's offense. These approaches locate the offense in the content, whether semantically or pragmatically conveyed. But we shall finally consider a non-content-based strategy, which locates the offensiveness of slurs somewhere else entirely—namely, in the fact that they are socially prohibited.

A first, simple proposal is a *semantic* one. Slurs have a different *literal content* from their neutral counterparts:

(1) Hank Aaron is an African American.
(2) Hank Aaron is a nigger.

On this proposal, the occurrence of the slur in (2) literally expresses more than "African American" in (1): perhaps "African American and despicable because of it". On this understanding, (1) is true, and (2) is simply false. And the additional, derogatory content in (2) helps explain its offense.

However there are reasons to think this proposal is not adequate. Suppose someone wants to deny whatever is ascribed to Aaron in (1). They can simply say

(3) Hank Aaron is not an African American.

(3) denies Aaron is African American: it is false, and inoffensive. Suppose someone wants to deny whatever is ascribed to Aaron in (2). They say:

(4) Hank Aaron is not a nigger.

On the semantic proposal, (4) denies that Aaron is African-American and despicable because of it: so it should be both inoffensive and true. But it is not true, or not obviously true (Anderson and Lepore, forthcoming, p. 6); and (4) is, moreover, offensive. This tells against the proposal. (Note there are theorists who disagree, saying the negation of a slurring sentence needn't be derogatory: Dummett, 1973; Hornsby, 2001; Hom, 2008.)

Considerations from indirect speech also undermine the semantic approach. Suppose John says "Aaron is African-American"; and Mark reports him by saying: "John said that Aaron is an African American". No problem: this is true, inoffensive, and follows the general guiding principle of reported speech that one should re-use expressions in reporting their first uses (Anderson and Lepore, forthcoming, p. 8). Suppose, though, Jim says "Aaron is a nigger": and Mark reports him by saying, "Jim said that Aaron is a nigger". That is just what Mark should say, on the semantic approach. Now, Jim said something offensive (and false, on the semantic approach): but Mark's report of Jim (whatever its truth value) is *also* offensive. And, interestingly, if we had nothing but Mark's report to go by, we might guess that Mark was a racist slurrer, and that Jim was *not*. A slurring indirect report may or may lay offense at the reportee's door: but it certainly lays offense at the reporter's door. We have some reasons, then, to doubt the semantic approach.

As an alternative, one might propose a *presuppositional* account of slurs. The behavior of slurs under negation, displayed above, provides a starting point here, since presuppositions seem to project through negation, as shown in the (unfortunate) classic (5) and (6):

(5) John stopped beating his wife.
(6) It's not the case that John stopped beating his wife.

Both statements presuppose "John beat his wife". A presuppositional approach might explain the behavior of slurs under denial and negation, since presuppositions project out of those contexts; and why occurrences of slurs in interrogative statements also tend to cause offense, as in—

(7) Are most landscapers spics?

—because presuppositions project out of interrogative contexts as well.

On the presuppositional proposal, the slur doesn't add the offensive content as part of the explicit truth conditions, but instead presupposes it. To return to our original examples, whilst (1) is true, (2) suffers from a presuppositional failure (Aaron is not despicable because of being African American). It therefore lacks a truth value, on a

Fregean understanding of presupposition; or it suffers whatever other shortcoming one takes presuppositional failure to give rise to. And on this account, (2)'s offensiveness is located in its presupposition.

However, indirect reporting presents a problem for this approach as well. Presuppositions and slurs behave differently in the complement clauses of "say": the presuppositionally encoded content of a complement clause following "say" is contained or isolated within the scope of the verb. But the offense of a slur in the complement clause projects out of the scope of "say", so that the offense is carried by the entire sentence. Notice the difference between (8) and (9):

(8) John said that Bill stopped beating his wife. But Bill never beat his wife.
(9) Maury said that spics are taking over Arizona. But Hispanics are not despicable because of being Hispanic.

Normally, "Bill stopped beating his wife" presupposes "Bill beat his wife". But in (8), the speaker does not present the latter as part of the common ground (Stalnaker, 2002): it's not assumed that hearers go along with it. "Bill beat his wife" is contained or isolated within the scope of "say". In (9), however, the offensiveness of "spic" is not contained or isolated within the scope of "say". If the offense were communicated presuppositionally, it wouldn't spread to the whole attribution: the reporter could explicitly distance himself from the reportee's presupposition, as in (8). But in the case of a slur, he can't. There is something infelicitous about following up the first statement in (9) with the second. So we have reason to doubt the presuppositional account.

A third strategy denies that the offensive content of slurs affects either their truth conditions or their presuppositions. A model here is Frege's understanding of "tone". For Frege, "this is my steed" and "this is my nag" have the same truth conditions and presuppositions, the difference between the two assertions being located elsewhere. One way of making this idea precise would draw on Grice's notion of *conventional implicature* (Grice, 1961, 1989). Consider:

(10) John is British and brave.
(11) John is British but brave.

For Grice, "but" and "and" are truth-conditionally equivalent, though "but" carries a notion of contrast that "and" lacks. The contrast is not a part of the asserted content of "but"; (11) is true as long as John is both British and brave. On this proposal, one might suggest that the difference between (1) and (2) is that "nigger" conventionally implicates something that "African American" does not. On this account, both are true, and the offense comes from the implicature. Since a conventional implicature cannot be cancelled, this explains why a speaker cannot felicitously follow up an utterance of (2) with, "But I don't mean to suggest anything offensive about Aaron!", any more than a speaker could follow up (11) with an utterance of—"and there's nothing surprising about a Briton being brave!". The proposal helps explain the behavior of slurs under denial and negation, since these do not affect the conventional implicature.

Despite its virtues, there are reasons to question the conventional implicature proposal. Numerous authors have expressed suspicion about the whole notion of conventional implicature (Bach, 1999; Carston, 2002, 2004). Putting that aside, this proposal does no better with indirect speech reports than does the presuppositional account.

Conventional implicatures are contained or isolated within the "says" complement just as much as presuppositions are. Consider:

(12) Pierre said that John is British but brave. However, there's nothing surprising about Britons being brave!

The speaker who utters (12) successfully insulates himself from anti-British prejudice; but the speaker of (9) does not successfully insulate himself from offensiveness towards Hispanics. So we must have reservations about this approach as well.

We have seen that there are problems with a semantic truth-conditional strategy, a presuppositional strategy, and a strategy drawing on conventional implicature. There is much more to be said, and there are arguments to which we cannot do justice here. (For vigorous defense of a semantic truth-conditional approach see Hom, 2008; for some important expressivist alternatives, see Kaplan, 1999; Richard, 2008.) But notice the three content-based approaches considered here also have a problem in common: a difficulty in accounting for the data with which we began, namely the varying degrees of offensiveness among slurs. What relevant difference in content—whether explicit, presupposed, or implicated—could account for the different degrees of offensive potency among slurs for different groups, or among co-referring slurs?

Given these difficulties, it may be worth looking in a completely different place for an understanding of the offensiveness of slurs: it may be worth looking at a *non-content-based* approach. Here is a simple alternative: slurs are *prohibited words* whose occurrences are offensive (Anderson and Lepore, forthcoming). This is a proposal not about the truth values of slurring sentences (being in principle compatible with a number of approaches to that question): rather, it is about the source of a slur's offense. *Prohibitionism* helps explain why offense projects out of the complement of "say" and quotation, why it is not challenged by denial, and why the speaker of a slurred word bears the offense, even in indirect reports. It helps explain the phenomenon of appropriation. The prohibition on a slur is not absolute. Like oil embargos or economic sanctions, prohibitions on slur use have exceptions, so that members of the targeted group may use them; and so that discussion in certain contexts, including academic contexts like this one, manage (we hope) to avoid offense. It helps explain how phonetic and orthographic accidents provoke offense, so that a word like "niggardly", though not semantically or etymologically related to the slur, may nonetheless be found offensive because the prohibition extends to words that sound like existing slurs (Kennedy, 2002). It might also explain how a slur may sometimes offend, even when insulated in quotation marks. To sum up then: whatever one's story about the content of slurs, whether semantically or pragmatically conveyed, that story has difficulty accounting for the offensive potency of the slur. The prohibitionist alternative offers an explanation worth taking seriously.

2 Speech Acts

Language using slurs and epithets, like the letter to Aaron, fits Mari Matsuda's description of hate speech, whose "message is of racial inferiority", is "directed against a historically oppressed group", and is "persecutory, hateful and degrading" (1993, 36). Matsuda sees hate speech as a kind of attack, and it includes, of course, speech that does not use slurs or epithets. Notwithstanding her talk of a "message", she is less interested in the

content of the words than in what is being *done* with these words (cf. Austin, 1962): hate speech is a speech act. The point of slurs or epithets, from this perspective, is not their content. But neither is it the simple fact that they are prohibited. It is that slurs and epithets are apt for performing certain speech acts: as Richard says, "what makes a word a slur is that it is used to *do* certain things, that it has . . . a certain illocutionary potential" (Richard, 2008, 1)—the potential to enable speakers to do things that e.g. "persecute" and "degrade'.

A speech act perspective is implicit in a United Nations condemnation of "propaganda" based on "ideas or theories of superiority of one race", that "attempts to justify or promote racial hatred and discrimination": speech that is an "incitement to, or act of, discrimination" or violence (UN, 1965). The UN focuses on speech addressed to prospective haters, rather than to targets of hate: hate speech *incites* and *promotes* racial discrimination, hatred and violence. Matsuda is interested in assault-like speech, the UN in propaganda-like speech, but both see hate speech as something that not only has certain effects (causing distress, spreading hatred), but also *is* something, in itself. In Austin's terms, both are talking about, *inter alia*, the "illocutionary" aspect of hate speech.

Austin distinguished the illocutionary aspect of an utterance from two others: the meaning, content, or "locutionary" act; and the effects on hearers and beyond, or "perlocutionary" act. Suppose (Austin's example) a speaker says "Shoot her!" There is content, a locutionary act: he said "Shoot" meaning "shoot with a gun", and "her", referring to the woman nearby. There are effects, a perlocutionary act: by saying "Shoot her!" he got his hearer to shoot the woman. But there is more: in saying "Shoot her", Smith *urged* the hearer to shoot. This is the illocutionary act, performed "in saying" those words.

In Austin's terms, hate speech can be an illocution that *persecutes* and *degrades*; or *incites* or *promotes* hatred, discrimination, violence. It is typically a more ordinary illocution as well: perhaps it *asserts* that there is a Jewish conspiracy, that hate will bring victory, that whites are superior; *orders* blacks to keep away. And it is typically a *perlocutionary* act that achieves effects, including changes in hearers" beliefs, emotions, and actions. But its illocutionary aspect is distinct from its locutionary content, and from its perlocutionary effects. Austin's distinction between illocutionary and perlocutionary acts offers a way to distinguish speech that *constitutes* racial oppression, and speech that *causes* racial oppression.

Hate speech is a kind of racially oppressive speech: letters "persecute" and "degrade" Aaron with assault-like hate speech; Nazi editorials "incite" and "promote" hatred against Jews with propaganda-like hate speech. But there may be other kinds of racially oppressive speech: a court says slaves are "incapable of performing civil acts", are "things, not persons"; a proprietor says "Whites Only". Speech like this is not, or not solely, assault-like or propaganda-like, nor does the label "hate speech" capture the core of what is done. Its point is to enact, or help enact, a system of racial oppression: it authoritatively *ranks* a certain group as inferior, *deprives* them of powers and rights, *legitimates* discrimination against them. Speech that does these things has, perhaps, the illocutionary force of *subordination* (cf. Langton, 1993, 2009). Note that its speakers are *authoritative*: a mere bystander who said "Whites Only" or "slaves . . . are things, not persons" would be doing something quite different. This fits an Austinian paradigm for illocutions that enact norms, or facts about permissibility: speaker authority is typically a "felicity condition" for their illocutionary success.

Drawing on Austin, we have sketched a distinction between three potentially overlapping classes of racially oppressive illocutions: assault-like hate speech, propaganda-like hate speech, and authoritative subordinating speech. More recent developments in pragmatics may shed further light on their workings.

Some have argued that a conversation is a game-like, rule-governed activity: players make moves, and the "score" or "common ground" of the conversation evolves in response to those moves (Lewis, 1979; Stalnaker, 2002). This notion of adjustment to "score" or "common ground" thus has significance not only in its role of tracking content, described in the previous section, but also, potentially, in a role of tracking illocution (Lewis, 1979). David Lewis compared conversation to a baseball game, viewing "score" as an abstract structure, consisting of propositions and norms, built up over time, constraining and enabling later moves available to players. But unlike a baseball game, conversational score tends to evolve in whatever way is required to make a move count as "correct play", following rules of *accommodation* (Lewis, 1979; Stalnaker, 2002; Atlas, 2004; von Fintel, 2008).

Presupposition introduction and Austinian performatives provide illustrations, for Lewis. If a speaker says, "Even George could win", then "straightway" the score or common ground is adjusted: she adds to the score not just "George could win", but also "George is an unpromising candidate". If an authoritative speaker says, "You are not permitted to cross the white line", then "straightway" the score is adjusted, and the bounds of permissibility are altered. In similar vein, Robert Stalnaker describes how the "common ground—the shared attitudes of conversational participants—evolves in response to what is said, following these rules of accommodation. (Lewis's picture of "moves" in a "language game", producing adjustments "straightway" to an abstract "score", exemplifies—arguably—Austin's non-causal picture of an illocutionary act. Stalnaker's picture is similar, but subtly different. His picture of adjustments to "common ground" in the attitudes of hearers exemplifies—arguably—Austin's causal picture of a perlocutionary act (Langton, forthcoming).)

The importance of this is that it enables us to see how racially oppressive speech acts might work in more informal contexts. Some kinds of racially oppressive speech acts (a "Whites Only" sign, an Alabama court pronouncement) fit a ceremonial Austinian paradigm, but many do not. Speakers often don't or couldn't pronounce, authoritatively and explicitly, "I hereby subordinate/persecute/discriminate". Speakers often don't intend to perform the relevant speech act, or use a performative tag ("I hereby'), or a relevant performative verb. They may lack formal authority, and may merely presuppose or implicate certain claims or norms. But this is no bar to their speech acts being illocutions that are racially oppressive (Langton and West, 1999; Langton, 2010). Thus propaganda-like hate speech may not explicitly assert that "Good Germans hate Jews," may not use a performative tag or illocutionary verb ("I hereby authorize and urge you to hate Jews"), but may still adjust the score, performing illocutions of authorizing and urging implicitly: so that it becomes appropriate for hearers to believe that good Germans hate Jews, and to hate Jews. Likewise assault-like hate speech may merely presuppose or implicate that its target is subhuman or animal-like, but the speaker may still succeed in placing that in the score. A restaurant proprietor may enact discriminatory policy, subtly placing it in the score, not by hanging a "Whites Only" sign but by encouraging staff to favour customers of a certain race ("Look at Mr. Wilson there—*just* the sort of customer we want").

Mary-Kate McGowan extends Austin's account on these lines, calling the relevant

phenomenon a "conversational exercitive" (2003, 2004). Focusing on the normative aspect of "score" or "common ground", she observes that conversational moves contributing to score are illocutions that alter normative facts about what is permissible, and even possible, in the conversation thereafter. Conversational exercitives work differently to paradigm Austinian speech acts. They work in covert ways: Speakers needn't intend to alter permissibility facts, needn't have special authority. In a conversation following "Even George could win," some moves are later impermissible ("Guess what, George is an unpromising candidate!"), some are more permissible (for example mockery of George). McGowan ultimately aims to explain how these local norms of a conversation hook up to wider social norms, so that enactment of a small racist local conversational norm can help constitute enactment of a larger racist social norm, set not by the law-enacting speech of authorities but by everyday speech of conversational speakers (2009).

One begins to see, then, how even an informal, conversational utterance could be an illocutionary act that helps constitute racial oppression, adjusting "score" and setting norms without being a grand ceremonial Austinian illocution; and how it could be a perlocutionary act that helps cause racial oppression, by helping cause racist beliefs and norms in its hearers, via adjustment to common ground.

The speech act perspective explored in this section provokes other questions. Perhaps it yields a special sense of *silencing*: If someone is unable to perform certain illocutions because of their race, might that be a kind of illocutionary disablement? (Langton, 1993) Perhaps a pragmatic story helps one see how rebellion is difficult, if *blocking* a presupposition is costly: must a blocker be uncooperative, derailing the conversation, going against "what we all take for granted'? (Langton and West, 1999) Perhaps the pragmatic story itself needs amending: If we are interested in how speech is used to transmit, not just knowledge, but hate, should "common ground" be extended to include attitudes "beyond belief", such as hate? (Langton, 2010) This look at racially oppressive speech through the lens of speech act theory has been brief but still suggestive, we hope, of fruitful possibilities.

3 Racial Generics

We have been considering utterances that are clearly racist: racial epithets or slurs, assault-like hate speech, propaganda-like hate speech, and authoritative subordinating speech. Now we turn to examples of speech that appear innocent, but may nevertheless undergird racial subordination.

Consider statements such as:

(15) Latinos are lazy.
(16) Muslims are violent.
(17) Blacks are sexually insatiable.

Those who make these sorts of claims often insist they are simply reporting the facts; however, counterexamples do not disrupt their confidence. Appiah, in considering this phenomenon, argues that racial prejudice just is this sort of epistemic failure. Racial prejudice is the disposition to hold beliefs—

> that are to the disadvantage of some race (or races) as opposed to others, and to

> do so even in the face of evidence and argument that should appropriately lead to giving those propositions up
>
> (Appiah, 1990, 15–16)

However, things are not so simple. Consider the following utterances:

(18) Ducks lay eggs.
(19) Lions have manes.
(20) Mosquitoes carry the West Nile Virus.

All seem to be true, even though one could provide many counterexamples of ducks that do not lay eggs (all male ducks), lions that don't have manes (all female lions), and mosquitoes that don't carry the West Nile Virus (over 99 percent, see Leslie, 2007, 376).

The utterances in both groups are generalizations, but they do not have the form, strictly speaking, of quantifications. Generalizations that omit quantifiers such as "some", "most", or "all" fall into the linguistic category of generics, and call for a different analysis than ordinary quantified statements. Importantly, for our purposes, generics need not be undermined by a counterexample, or even many counterexamples.

Although there is no consensus on the best account of generics, Sarah-Jane Leslie suggests convincingly that generics are the expression of a very primitive "default mode of generalizing," that picks up on significant or striking properties and links them to a psychologically salient kind.

In exploring this, we must be sensitive to different sorts of generalizations. Prasada and Dillingham (2006, 2009) highlight two sorts. *Principled* generalizations involve properties determined by the kind of thing something is (for example having four legs, for a dog). *Statistical* generalizations involve properties not determined by the kind of thing something is but are highly prevalent connections to the kind (for example cars having radios). Principled generalizations support formal explanations (Fido has four legs because he is a dog), normative expectations (Fido should have four legs, is defective if not), and expectations of prevalence (most dogs have four legs) (Leslie et al., 2009, 479).

Leslie notes that this distinction between principled and statistical generalizations, although important, is not sufficient, by itself, to accommodate the range of different generics, for, as noted above, there are true generics that do not ascribe a prevalent property to the kind, such as:

(18) Ducks lay eggs.

Leslie proposes instead not two, but three, classes (Leslie, 2008, 43). First, there are *characteristic generics*, which correspond roughly to Prasada and Dillingham's "principled" generics. Cases such as

(21) Tigers have stripes.

and

(22) Dogs have four legs.

These assert more than a statistical correlation between tigers and stripes, or dogs and four legs. They purport to tell us what is characteristic of the kind, or what a good example of the kind will exhibit. And if we consider what is characteristic of the *kind*, rather than of each individual, we can accommodate

(18) Ducks lay eggs.

even if the majority of ducks don't lay eggs. We have background knowledge that—

> provides an outline of information to be gathered about a new kind; characteristic dimensions provide a learner with an informational template. When a value is found for a characteristic dimension of a kind, it is hereby generalized to the kind by the basic generalization mechanism, and so the generic that predicates that property of the kind is accepted. Ducks, being an animal kind, have reproduction as a characteristic dimension, so the inductive learner looks for a value to fill the dimension; even limited experience will deliver *laying eggs* as the appropriate value, and so the property is generalized to the kind and "ducks lay eggs" is accepted as true.
>
> (Leslie, 2008, 32–3)

Second, there are *striking property generics*. This extends Prasada and Dillingham's "principled" generics to cases where the property is not prevalent. To accommodate

(20) Mosquitoes carry the West Nile virus.

even though only a small fraction do, Leslie says: "if the property in question is the sort of property of which one would be well served to be forewarned, even if there were only a small chance of encountering it, then generic attributions of the property are intuitively true (2008, 15)".

On her account, being a member of the kind must be a reasonably good predictor of the striking property; members that don't have the property must be disposed, under the right circumstances, to have it (2008, 41).

Third, there are *majority generics*. This corresponds to Prasada and Dillingham's "statistical" generics. Speakers are willing to assent to weaker generics such as

(23) Cars have radios.
(24) Barns are red.

that capture only statistical or enumerative generalizations. However speakers do find these less natural than characteristic generics, and resist alternative syntactic forms for the generic. Thus in the case of the indefinite singular form—

(25) A tiger has stripes.

—is considered more natural than

(26) A barn is red.

(See Leslie et al., 2009, 480, 482, 484.)

Leslie's account of generics is controversial, and more complex than suggested here (see Leslie, 2008, 35). The message for our purposes, though, is simple enough. Generics are a distinctive kind of statement that should not be treated as ordinary quantified statements: They draw heavily on background knowledge and patterns of inference to highlight a significant property of a kind—a property which may be characteristic, striking, or just statistically correlated with the kind.

Leslie is right to complicate Prasada and Dillingham's view by allowing for striking property generics in which the property attributed is not prevalent among members of the kind. However, the distinction between principled and statistical connections still seems important. In principled connections there is something about being a member of the kind *as such* that explains why those members that have the property do so. Prasada and Dillingham's mistake was to assume that when a principled connection holds between members of a kind and some property, that property is prevalent among members of the kind. But it may not be prevalent because the circumstances prevent members of the kind from manifesting the property, or because it is typical of the kind that only certain members manifest it. The crucial fact is this: Those members which do have the property are disposed to have it by virtue of the fact that they are members of the kind. If one goes in for essentialist talk, one might say that having the property is grounded in or flows from the nature or essence of the kind. Laying eggs is grounded in the nature of ducks (being oviparous is part of what ducks are). Carrying the West Nile Virus is something mosquitoes are disposed to do by virtue of being mosquitoes. It does not follow, however, that all or most members of the kind have the property.

If we adjust the idea of principled connections in this way, Leslie's characteristic and striking property generics both turn out to be subcategories of principled generics. Such generics entail, or implicate, claims about natures or essences. Although the utterance of a majority generic such as:

(23) Cars have radios.

does not imply that it flows from the nature of cars that my VW is disposed to have a radio, the utterance of what appears to be a characteristic generic such as—

(15) Latinos are lazy.

—does imply that laziness flows from the nature of Latinos, as such.

This helps explain why some speakers refuse to withdraw racially stereotyping claims when confronted with counterexamples. Even if many Latinos do not demonstrate laziness, that doesn't show it isn't part of their nature. They may just be "defective" Latinos; they may have been denied the opportunity to express their nature (but it could happen at any time!). In order to undermine the characteristic generic, one must go beyond counterexamples and argue against the modal/explanatory claim.

Given the different generics that might be expressed by a bare plural, it is unsurprising that what they communicate is slippery. Two kinds of slip are highly relevant to the issue of race: (i) a slippage between characteristic and striking property generics, and (ii) a slippage between characteristic and majority generics.

To explore the first slippage between characteristic and striking property generics, consider an ethnic generic such as:

(16) Muslims are violent.

It is tempting to regard (16) as a striking property generic which, like

(20) Mosquitoes carry the West Nile Virus.

is assertable based on a small number of instances. However, given that the same statement form can also be used to assert a characteristic generic, there is an easy slide from accepting (16) as a striking property generic to the conclusion that it is characteristic of Muslims to be violent.

> The basic idea is quite simple: just as it takes but a few instances of sharks attacking bathers, or of mosquitoes carrying the West Nile virus, for us to make the corresponding category-wide generalization, so also a strikingly negative action on behalf of a few members of a racial minority may lead us to a general belief concerning their entire racial category. A belief in the essentialized nature of racial categories makes them appear to be appropriate loci of striking property generalizations; our disposition to thoughtlessly generalize dangerous or harmful information does the rest . . . Extreme and aberrant actions on behalf of the few can thus lead to conclusions concerning all.
>
> (Leslie, forthcoming)

Psychological research shows, moreover, that striking property generics are quickly accepted based on few instances, whereas characteristic generics, once accepted, are resistant to refutation. Given the slippage between these two forms, properties not indicative of a general pattern get attributed to a group and then stick (Cimpian, Brandone, and Gelman, 2010). Easy to accept, hard to refute is not an epistemically or politically promising combination.

The second kind of slippage occurs between characteristic and majority generics. Without taking a stand on the semantics of generics, we offer the following pragmatic proposal. The utterance of a generic that is not obviously just a majority generic implicates that the property ascribed is grounded in the nature or essence of the kind (Haslanger, 2010). The implicature can be canceled explicitly or implicitly by context, but otherwise the implicature enters the common ground of the conversation (Stalnaker, 2002).

Attending to this slippage between majority and characteristic readings helps us explain certain shifts in the defense of generics. Consider again:

(15) Latinos are lazy.

Does (15) assert a majority generic or a characteristic generic? Interpret (15) as a majority generic. To combat it, one provides many counterexamples. However, the speaker can then suggest that, although many Latinos aren't lazy, they tend to be—thus embracing the characteristic generic. Instead interpret (15) as a characteristic generic. To combat it one provides evidence that, say, Latinos show no greater tendency towards laziness than any other group. The speaker can then suggest that, although it is not part of the nature or essence of Latinos to be lazy, most are. This slide back and forth between different interpretations of the utterance allows speakers to avoid taking responsibility for the implications of their claims.

What are the morals of the story? We suggest three.

First moral: If a characteristic racial generic is asserted, it should be rejected, because is false, and also politically problematic; it presents social artifact as racial essence, and by masking oppression is a barrier to social change. Besides masking, it also assists oppression via social looping effects: If individuals are made to become as they are assumed to be by nature, then statistical generics are often true, and their truth seems to confirm the faulty assumption we started with.

Second moral: If a striking property racial generic is asserted, it should also be rejected because it is expressed in a statement form that is too readily interpretable as a characteristic generic.

Third moral: If a statistical or majority racial generic is asserted it should be rejected not only if it is false, but also, surprisingly, if it is true—as it might sometimes be, given that majority generics are no stronger than a quantified "most" claim. A true statistical or racial generic should, usually, be rejected metalinguistically (Horn, 1985) because it implicates a characteristic generic. Metalinguistic rejection of the statistical or majority generic, like the rejection of the characteristic generic, blocks the harmful invocation of racial natures. A much better approach to true statistical generics is to insist on explicitly quantified statements instead ("Most . . ." or "Some . . .") which avoid slippage to the characteristic generic.

To sum up: generics may come naturally to us, but their assertion in racial discourse can do much epistemic and moral harm.

Bibliography

Altman, A. 1993. "Liberalism and Campus Hate Speech: a Philosophical Reflection," *Ethics* 103: 302–17.

Anderson, L. and Lepore, E. Forthcoming. "Slurring Words," *Noûs*.

Andreasen, R. 1998. "A New Perspective on the Race Debate," *British Journal of the Philosophy of Science* 49(2): 199–225.

Appiah, K. A. 1990. "Racismsin David Theo Goldberg," (ed.), *Anatomy of Racism*, Minneapolis: University of Minnesota Press.

—— 1996. "Race, Culture, Identity: Misunderstood Connections," in A. Appiah and A. Gutmann (eds.), *Color Conscious*, Princeton, NJ: Princeton University Press.

Atlas, J. D. 2004. "Presupposition," in L. R. Horn and G. L. Ward (eds.), *The Handbook of Pragmatics*, Oxford: Blackwell. 29–52.

Austin, J. L. 1962. *How to Do Things with Words*. London: Oxford University Press.

Bach, K. 1994. "Conversational Impliciture," *Mind and Language* 9(2): 124–162.

—— 1999. "The Myth of Conventional Implicatures," *Linguistics and Philosophy*, 22.4 (Aug1999): 327–366.

Bianchi, C. 2008. "Indexicals, Speech Acts and Pornography," *Analysis* 68: 310–316.

Bird, A. 2002. "Illocutionary Silencing," *Pacific Philosophical Quarterly* 83: 1–15.

Brontsema, R. 2003. "A Queer Revolution: Reconceptualizing the Debate over Linguistics Reclamation," *Colorado Research in Linguistics* 17, (1): 1–16.

Butler, J. 1997. *Excitable Speech: A Politics of the Performative*. New York and London: Routledge.

Carston, R. 2002. *Thoughts and Utterances*. Oxford: Basil Blackwell.

—— 2004. "Relevance theory and the saying/implicating distinction," in L. Horn and G. Ward (eds.) *Handbook of Pragmatics*. Oxford: Blackwell 607–632.

Catterall, H. T. (ed.) 1926. *Judicial Cases Concerning Slavery and the Negro*. Washington, D.C.: Carnegie Institute.

Cimpian, A., Brandone, A. C. and Gelman, S. A. 2010. "Generic Statements Require Little Evidence for Acceptance but Have Powerful Implications." *Cognitive Science* 34(8): 1452–1482.
Delgado, R. 1993. "Words that Wound: A Tort Action for Racial Insults, Epithets and Name Calling," in K.W. Crenshaw et al. (eds.), *Words that Wound*. Boulder, CO: Westview Press.
Dubois, W. E. B. 1996. "The Name "Negro"" in E. J. Sundquist (ed.) *The Oxford W.E.B. Dubois Reader*. Oxford: Oxford University Press 70–72.
Dummett, M. 1973. *Frege's Philosophy of Language*. Oxford: Clarendon Press.
Feinberg, J. 1985. *Offense to Others New York*. London: Oxford University Press.
Fiss, O. 1996. *The Irony of Free Speech*. Cambridge, MA: Harvard University Press.
Frege, G. 1918/19. "The Thought: A Logical Inquiry," in R. Harnish (ed.) *Basic Topics in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall 517–535.
Fricker, M. 1998. "Rational Authority and Social Power: Towards a Truly Social Epistemology," *Proceedings of the Aristotelian Society*, 98(2): 159–177.
—— 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.
Glasgow, J. 2009. *A Theory of Race*. New York: Routledge.
Grice, H.P. 1961. "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, Supp. xxx: 121–153.
Grice, P. "Logic and conversation," in *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
Habermas, J. 1984. *The Theory of Communicative Action*, translated by Thomas McCarthy. Boston: Beacon Press.
Haslanger, S. 2008. "A Social Constructionist Analysis of Race," in B. A. Koenig, S. Soo-Jin Lee and S. Richardson (eds.), *Revisiting Race in a Genomic Age*, New Brunswick: Rutgers University Press 56–69.
—— 2010. "Language, Politics and 'the Folk': Looking for 'The Meaning' of 'Race'". *The Monist*93:169–187.
Hom, C. 2008. "The Semantics of Racial Epithets," *Journal of Philosophy* 105: 416–440.
Horn, L. 1989. *A Natural History of Negation*. Chicago: Chicago University Press.
Hornsby, J. 1994. "Illocution and its Significance," in S.L. Tsohatzidis (ed.) *Foundations of Speech Act Theory*, London: Routledge.
—— 1995. "Disempowered Speech," *Philosophical Topics* 23: 127–147.
—— 2001. "Meaning and Uselessness: How to Think about Derogatory Words," in P. French, et al. (eds.) *Midwest Studies in Philosophy XXV*: Blackwell 128–141.
James, M. 2008. "Race," in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), http://plato.stanford.edu/archives/fall2008/entries/race/
Kaplan, D. 1999. "The Meaning of "Ouch" and "Oops": Explorations in the Theory of Meaning as Use," ms. UCLA.
Kennedy, R. 1999–2000. "'Nigger'? and Other Considerations," *The Journal of Blacks in Higher Education* 26: 86–96.
—— 2003. *Nigger: The Strange Career of a Troublesome Word*, New York: Vintage Press.
King, M. L. 1963. "I have a Dream" speech, MLK Online http://www.mlkonline.net/ Accessed June 8, 2010.
Kitcher, R. 1999. "Race, Ethnicity, Biology, Culture," in L. Harris (ed.) *Racism*, New York: HumanityBooks.
Langton, R. 1993. "Speech Acts and Unspeakable Acts," *Philosophy and Public Affairs* 22: 305–330.
—— Forthcoming. "Beyond Belief: Pragmatics in Hate Speech and Pornography," in M.K. McGowan and I. Maitra (eds.) *What Speech Does*, Oxford: Oxford University Press.
—— 2009. *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. Oxford: Oxford University Press.
Langton, R. and Hornsby, J. 1998. "Free Speech and Illocution," *Journal of Legal Theory* 4: 21–37.
Langton, R. and West, C. 1999. "Scorekeeping in a Pornographic Language Game," *Australasian Journal of Philosophy* 77: 303–19.
Leslie, S. J. 2007. "Generics and the Structure of the Mind," Philosophical Perspectives 21 (1): 375.
—— 2008. "Generics: Cognition and Acquisition," *The Philosophical Review* 117(1): 1.
—— Forthcoming. "The Original Sin of Cognition: Fear, Prejudice and Generalization." *Journal of Philosophy*.
Leslie, S. J., Khemlani, S., Prasada, S. and Glucksberg, S. 2009. "Conceptual and Linguistic Distinctions between Singular and Plural Generics," Proceedings of the 31st Annual Cognitive Science Society. Amsterdam: Cognitive Science Society.

Lewis, D. 1978. "Truth in Fiction," *American Philosophical Quarterly* 15, 37–46, reprinted in Collected Papers, Vol. I, 261–280.
—— 1979. "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8: 339–59, reprinted in Philosophical Papers, Vol. I, Oxford: Oxford University Press, 1983, 233–249.
MacKinnon, C. 1987. *Feminism Unmodified*. Cambridge, MA: Harvard University Press.
——1993. *Only Words*. Cambridge, MA: Harvard University Press.
Matsuda, M. with Lawrence, C.R. III, Delgado, R. and Williams Crenshaw, K. 1993. *Words that Wound: Critical Race Theory, Assaultive Speech and the First Amendment*. Boulder, CO: Westview Press.
McGowan, M. K. 1993. "Conversational Exercitives and the Force of Pornography," *Philosophy and Public Affairs* 31: 155–189.
—— 2004. "Conversational Exercitives: Something Else We Do With Our Word" *Linguistics and Philosophy* 27: 93–111.
—— 2009. "Oppressive Speech," *Australasian Journal of Philosophy* 87: 389–407.
—— and Ishani, M. 2007. "Limits of Free Speech: Pornography and the Question of Coverage," *Legal Theory*, 13: 41–68.
Omi, M. and Winant, H. 1994. *Racial Formation in the United States*. New York: Routledge.
Outlaw, L. 1990. "Toward a Critical Theory of Race," in D.T. Goldberg (ed.) *Anatomy of Racism*, Minneapolis: University of Minnesota Press.
Potts, C. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
Prasada, S. and Dillingham, E. M. 2006. "Principled and Statistical Connections in Common Sense Conception," *Cognition* 99(1): 73–112.
—— 2009. "Representation of Principled Connections: A Window onto the Formal Aspect of Common Sense Conception," *Cognitive Science* 33(3): 401–448.
Richard, M. 2008. *When Truth Gives Out*. Oxford: Oxford University Press.
Rosengren, J. 2008. *Hammerin' Hank, George Almighty and the Say Hey Kid*. Naper, IL: Sourcebooks.
Saul, J. M. 2008. "Pornography, Speech Acts and Context," *Proceedings of the Aristotelian Society*, 229–248.
Streicher, J. (ed.). 1943. *Der Stürmer* (Streicher Verlag), from an essay by Ernst Hiemer, "The Holy Hate"; German Propaganda Archive, translated and edited by Randall Bytwerk, http://www.calvin.edu/academic/cas/gpa/ww2era.htm Accessed June 3, 2009
Schwartzman, L. 2002. "Hate Speech, Illocution, and Social Context: A Critique of Judith Butler," *Journal of Social Philosophy* 33: 421–41.
Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
—— 1979. *Expression and Meaning*. Cambridge: Cambridge University Press.
Sperber, D., Premack D. and Premack A. J. (eds.). 1994. *Causal Cognition: A Multidisciplinary Approach*. Oxford: Clarendon Press.
Stalnaker, R. 2002. "Common Ground," *Linguistics and Philosophy* 25(5): 701–721.
Strawson, P.F. 1964. "Intention and Convention in Speech Acts," *Philosophical Review* 73: 439–460.
United Nations (UN) 1965. International Convention on the Elimination of All Forms of Racial Discrimination, Article 4, http://www.hrcr.org/docs/CERD/cerd3.html Accessed June 3, 2009.
von Fintel, K. 2008. "What is Presupposition Accommodation, Again?" *Philosophical Perspectives* 22: 137–170.
Wieland, N. 2007 "Linguistic Authority and Convention in a Speech Act Analysis of Pornography," *Australasian Journal of Philosophy* 85: 435–56.
Williamson, T. 2007. "Reference, Inference, and the Semantics of Pejoratives," in J. Almog and P. Leonardi (eds.)*Festschrift for David Kaplan*, Oxford University Press: Oxford.
Zack, N. 2002. *Philosophy of Science and Race*. New York: Routledge.

6.5
APRIORITY

Sinan Dogramaci

Introduction

What Is Apriority?

Apriority characterizes a way our beliefs acquire justification. Justification acquired purely in this way is apriori justification, as opposed to empirical (also called "aposteriori") justification. Apriority then derivatively applies to knowledge or to propositions when apriori justification is or could be had, respectively. In English, philosophers today may use either the single or double words, "apriori" or "a priori," sometimes italicizing "*a priori*" to indicate use of the original Latin.

When does a belief have apriori justification? The traditional answer is: when its epistemic justification is purely acquired independently of experience. However, that gloss stands in need of several refinements, and there is no consensus as to how to make them. First, what is *epistemic* justification? Roughly, it's justification had when a person's belief is *likely* from her perspective to be true. It's unclear, though, how to fully explicate the relevant notion of likelihood in noncircular terms. Second, what counts as experience here? Perceptual experience, the hallmark of empirical justification, certainly counts, but whether, for instance, introspection or other modes of self-reflection should also count is unclear. Finally, and perhaps most famously, no beliefs, much less justified ones, can be had *totally* independently of experience. As Kant famously put it, "how else should the cognitive faculty be awakened into exercise if not through objects that stimulate our senses and in part themselves produce representations . . . ?" (Kant (1781/1998): B1) A belief's apriori justification must somehow hold not *in virtue of* any experiences, even if some experiences are causal prerequisites.

Many important philosophers, especially in recent history, have questioned whether the notion of apriori justification is coherent, has theoretical utility, or has interesting scope (Quine (1951); Devitt (1998); Harman (2003); Stalnaker (2003); Williamson (2008).). These challenges shouldn't be dismissed, but this won't be where we face up to them. I'll assume we have a working grasp of a coherent notion of the apriori and can rely on a working hypothesis that it is theoretically sound. Any investigation into the apriori proceeds with a hope to refine and a willingness to revise our initial assumptions.

What is a discussion of a patently epistemological notion doing in a volume on the philosophy of language? The basic idea of an intimate connection between apriority and *analyticity* is centuries old. In the last twenty or so years, though, an interesting connection has been seen between apriority and *conceptual role semantics*. After introducing a little more context, my main aim will be to critically examine that latter connection.

The Traditional Scope of the Apriori

Apriority has been alleged to characterize propositions across a diverse range of topics. A sample of such topics includes: morality, color relations, metaphysical modality, axioms of recently developed formal theories such as set theory and probability theory, and philosophy itself. Controversial as the notion of apriority itself already is, its application to any particular one of these topics is yet a further controversy. However, *if* the notion of apriority is in good standing, then there are two core topics to which it most uncontroversially applies: *mathematics* and *logic*. If these topics aren't apriori, apriority is of little interest.

All of mathematics is generally taken to be apriori, but the propositions whose apriority is most discussed tend to be claims of arithmetic, in particular the most elementary of such truths. Consider your justification for believing that every number has a successor. This isn't something you plausibly know on the basis of any perceptual experience. This is a paradigmatic example of apriori justification.

Within logic broadly understood, we need to distinguish two areas of apriority. First, we have apriori justification for believing formal *meta*logical laws. For example, we're justified in believing in the validity of Modus Ponens, i.e. in the claim that every argument of the form *p*, *if p then q*, *therefore q* is truth-preserving. By contrast, there is our apriori justification for believing and inferring logical truths. When you reason your way toward some novel logical truth, reasoning *by* Modus Ponens and other such rules, apriori justification is transmitted from premises to conclusion in each inferential step. In some cases involving zero premises, justification can be generated, not merely transmitted. (An example of a rule that doesn't simply transmit but generates apriori justification might be the rule that licenses belief in the logical truth that either alien life exists or it doesn't.)

This generation and transmission of apriori justification within logical reasoning is of special importance. While it's been argued that our justification in some of these core mathematical, metalogical, and logical areas reduces to our justification in others, the one area where our justification seems *least* susceptible to reduction is this last area of logical reasoning. (For discussion of such reductions, see Dummett (1973), Boghossian (2001), Wright (2001) and Dogramaci (2010) on logical reasoning and metalogical laws, and see Hale and Wright (2001) on mathematics and logic.) For that reason, and because space is limited, in what follows I will primarily focus on apriority in cases where we reason toward the logical consequences of what we know.

Two Main Views of What Explains Apriority

In the contemporary debate, there is something of a competition between two main views about what explains apriori ways of acquiring justification. Though no labels are completely standard, I'll call these *intuition-based views* and *meaning-based views*. The latter views are our focus here, but it will be useful to briefly talk about intuition-based views. The ultimate evaluation of meaning-based views largely comes down to how their net plausibility compares with that of intuition-based views.

An intuition-based view aims to explain the apriori way a person acquires justification by appealing to intuitions she has, a kind of mental state modeled on perceptual experience. Such a quasi-perceptual state of intuition was perhaps most famously invoked by Descartes, in the notorious form of clear and distinct perception. Intuition-based

views have, in the last twenty years or so, again come into fashion (BonJour (1998); Huemer (2007); Sosa (2007); Bengson (2010); Chudnoff (forthcoming); Bealer (2000) is a hybrid intuition- and meaning-based view); at the same time, naturally, criticism of intuition-based views has also seen a rise (e.g. Weinberg et al. (2001)).

Perhaps the most attractive prospect that intuition-based theories hold is that of a fully general and unified explanatory theory of epistemic justification: as perceptual experiences explain empirical justification, intuitions explain apriori justification. To secure this advantage though, the intuition-theorist cannot construe intuitions as anything too obscure (such as clear and distinct perception). Intuitions should be a kind of mental state that's very similar to perceptual experiences. As such, many contemporary theorists (e.g. Bealer, Huemer, Chudnoff) construe intuitions merely as *seemings*, the mental state you're in whenever something seems to you to be the case. On such views, just as your belief that blood is red is justified by its visually seeming that blood is red, your belief that either aliens exist or they don't is justified by its seeming so, not perceptually but merely "intellectually".

Two major challenges shape the current debate over intuition-based views. One of these concerns how intuitions help explain *inferential* justification. Even if we suppose an intuition justifies belief in a metalogical claim that, say, a certain form of inference is valid, there's a question about whether that could explain how justification gets transmitted across beliefs in an inference. The worry is inspired by a famous dialogue by Lewis Carroll (Carroll (1895)). Carroll's cryptic dialogue appears to raise a worry that, if inferential justification involves knowledge of validity, then some kind of infinite regress ensues. In a recent sequence of articles, Paul Boghossian has reinvented the challenge not as a threat of regress, but as a charge that knowledge of validity is simply explanatorily redundant when it comes to explaining certain cases of inferential justification. I *already know* the conditional *if p then, q;* it adds nothing explanatorily significant for me to have an *intuition* that (necessarily) *if: p and if p then q, then: q.* (Boghossian (2001); Boghossian (2003). Wright (2001) critically examines Boghossian's challenge.)

The other challenge to intuition-based views can be posed in the form of a dilemma. On the one hand, intuitions had better not be anything supernatural; it's not plausible that any faculty of rational insight puts us in, say, causal touch with some heavenly realm of mathematics or logic. This pressures us toward a nonexotic construal of intuitions, something like mere seemings. On the other hand, if intuitions are constituted just by a seeming, an internal mental state that anyone could have at anytime, then what reason is there to think intuitions are *reliable?* We can tell a plausible causal story about how perceptual experiences are reliable, but a causal story will not do for intuitions. The challenge here was first posed by Benacerraf (1973), and later sharpened by Field (1989) as follows: if you think there is no explanation of how your beliefs about some topic are reliable, then you are rationally obliged to give those beliefs up.

Meaning-based views of apriority enter the debate advertising themselves as especially well suited to explain both reliability and inferential justification. What, then, is a meaning-based view?

The starting point or background premise of any meaning-based view is some form of conceptual role semantics, henceforth "CRS." Different meaning-based theorists accept different forms of CRS. The weakest thesis common to all meaning-based views is that a person's disposition to engage in certain patterns of reasoning is a necessary condition on her understanding certain meanings or grasping certain concepts. Note, this is stronger than just a claim about what's required to have representational states or even

to have beliefs: the claim is about what's required to have states with certain contents, contents involving particular concepts. Note, though, that the claim is weaker than the common *reductive* forms of CRS: meaning-based theorists needn't accept that there are non-intentionally describable dispositions constituting our intentional mental states. (Boghossian and Peacocke, prominent defenders of meaning-based views, have both rejected reductive forms of CRS (Boghossian (1989); Peacocke (1998)). For an extensive critique of CRS in all forms, see Williamson (2008).)

Now, a conceptual role semantics of any kind for any of our words or concepts is controversial. CRS is perhaps on firmest ground, when it comes to our understanding of logical constants. It certainly seems impossible that a subject fully understands the standard logical connectives while lacking any disposition to draw the inferences governed by certain canonical rules. For example, suppose we want to know whether this linguistic (or mental) symbol in a person's language (of thought), •, expresses the concept of *conjunction*. An extremely intuitive necessary (even sufficient) condition on that is that she is disposed to infer $p \bullet q$ from the pair of beliefs p and q, and is disposed to infer p, and to infer q, from the belief $p \bullet q$. Similarly, anyone who systematically refused to infer q from the pair of beliefs p and $p * q$ would not plausibly mean the conditional by that connective.

(Could someone grasp a concept, by having a disposition that she never exercises? We can allow for such a bizarre possibility. The person succeeds in understanding $p * q$ as a conditional, because if she *were* also to believe p and to consider whether q, she *would* infer q. Since she never gets the opportunity to learn p, she never makes the inference. This reasoner might even be perfectly rational, though she might not get to form as many beliefs as she'd like.)

Supposing a CRS for logical concepts is roughly right, then meaning-based views may begin to look attractive. What we notice is an apparent *coincidence*, literally speaking, between the inferences crucial to understanding logical connectives and the inferences that transmit justification in an apriori way. But we shouldn't rush to conclusions. CRS itself has no immediate epistemological implications.

The claim of interest, and the claim that defines something as a meaning-based view, is that the conditions on understanding a meaning or grasping a concept somehow *explain* the apriori way that beliefs acquire justification. What, then, is the explanatory link from CRS to justification supposed to be? And, how does the link address the challenges that faced intuition-based views? We'll now look at the leading contemporary proposals, and the problems those proposals face.

Meaning-based Explanations of Reliable Inference

Connecting Meaning and Reliable Inference

Each of the leading contemporary advocates of meaning-based views, namely Peacocke, Boghossian, Bealer, and Hale and Wright, have, at one time or another, endorsed the following powerful claim: When a disposition to engage in certain reasoning is essential to understanding meaning, this secures that reasoning's reliability. To see what can be said in defense of this claim, we have to step back from the issue of apriority for a moment.

Supposing CRS for the logical connectives is roughly right, there is a question that arises before we even bring up apriority at all. What explains the connection between certain inferences and certain concepts? In other words, why are *these* inferences crucial

to grasping conjunction; why are *those* inferences crucial to grasping the conditional? There must be some general theory dictating how concepts get paired with their associated conceptual roles.

A clue to answering that question emerges when we consider certain defective conceptual roles. Prior (1960) concocted the following famous pair of inference rules, alleging they define a connective he dubbed *tonk:* from *p,* infer *p*-tonk-*q*; and from *p*-tonk-*q,* infer *q*. Prior's intent was to (light-heartedly) mock the claim of a connection between an inference's reliability and its being essential to understanding the meaning of a connective; *tonk's* inferences, far from being reliable, lead to any absurdity. But a different and very plausible moral was drawn, first by Belnap (1962) and by many others since, namely that there simply is no genuine logical operation identified by the connective *tonk; tonk* is simply meaningless. That, in turn, suggested there must be substantial constraints on when an inference is essential to understanding a meaning, at least one of those constraints being that the inferences are *truth-preserving*. A truth-preservation constraint would also begin to explain the pairings between the standard logical connectives and their canonical inferences: Those connectives each make their associated inferences truth-preserving. For example, if • expressed disjunction or the conditional, rather than conjunction, then the inference to *p* (or to *q)* from *p* • *q* would be highly unreliable.

If the general theory dictating which conceptual roles pair with which concepts does include such a constraint, then the meaning-based theorist is a step closer to explaining how apriority ensures reliability. The full explanation isn't in hand yet, since we still need the part of the explanation that says *why* non-truth-preserving inferences cannot occur in the conceptual roles essential to understanding meanings. If we leave it a pure coincidence that those inferences are truth-preserving, then the position looks no better than that of the intuition-theorist who simply asserts a coincidence between intuitions and truths. The Benacerraf–Field challenge to our ability to rationally maintain our beliefs in this area would remain.

Now, there is a plausible idea the meaning-based theorists can draw on to explain the connection between meaning and reliability. The idea is that in coming to understand the meaning of a new word or grasp a new concept, we *implicitly stipulate* that this word/concept expresses *whatever semantic value makes certain inferences truth-preserving*. For example, we may fix the meaning of • as conjunction, in part, by implicitly stipulating that • expresses whatever truth-function makes certain inferences truth-preserving (namely, the canonical inferences for conjunction mentioned earlier). The inference rules for *tonk* then don't help to fix any meaning, because the relevant stipulation cannot be fulfilled.

Much more must be said, of course, to fully develop and defend this sort of strategy for connecting meaning and reliability. Christopher Peacocke has done extensive work on this project, engaging with it in many influential works (including Peacocke (1993); Peacocke (2000); Peacocke (2004)). Boghossian (1996), Bealer (2000), and Hale and Wright (2001) also defend the strategy, though all in different ways. Against all these views, Paul Horwich argues that stipulations of truth cannot in any way explain how we understand meanings, nor can facts about meaning-constitution secure the reliability of our inferences (Horwich (2005): chapter 6).

The outcome of this debate is important: it decides whether meaning-based views really are in a superior position to address the Benacerraf–Field challenge. However, even if that challenge were put to rest, there remain serious doubts about whether a

meaning-reliability connection could explain everything we need to explain about apriority. I turn to those doubts now.

Doubts about Whether a Meaning-Reliability Connection Fully Explains Apriority

Apriority, the thing we want a theory to explain, is a way of acquiring epistemic *justification*. Now, some philosophers, so-called "reliabilists," hold the view that the one and only way for a belief to acquire epistemic justification just is that it be the product of a reliable belief-forming process. But, that view is highly controversial, and in any case, all of the leading meaning-based theorists explicitly reject it. Two of them, though, Peacocke and Bealer, do claim that a *guarantee* of reliability would explain apriori justification. Here is Peacocke's position:

> According to the metasemantic theory [Peacocke's theory], in using an a priori way of coming to judge that *p* a thinker is using a method which guarantees, as a result of the very nature of *p* and the way in which its truth-condition is determined, that the thinker judges that *p* only if it is the case that *p*. When the soundness of a method is thus internally related to what it is for the content to be true, it is hard to see what more could be required for knowledge. Such a constitutive grounding of the soundness of the method goes far beyond merely reliabilist conditions for knowledge. (Peacocke (2004): 173)

(The view is also put forward in Peacocke (1993): 190; Peacocke (2000): 265. For a similar view, see Bealer (2000): 7–10. Bealer doesn't demand a guarantee, only a "strong modal tie between [a method's] deliverances and the truth" (p.9). For brevity, I'll stick to using "guarantee", but the following comments apply equally if we substitute "strong modal tie").

In the passage, Peacocke mentions conditions for knowledge, but I want to set knowledge aside and focus on the weaker condition of being epistemically justified. Should we agree with the suggestion that if a way of forming a belief is guaranteed to be reliable, that *explains why* it is epistemically justified? I don't know any simple objection that would compel a staunch advocate like Peacocke or Bealer to concede this isn't a sufficient condition for both knowledge and justification. However, when they offer the condition as an *explanatory* one, as they do, their view does face pressure: there are two heavy burdens the view has to shed before it can be seen as part of the best overall explanatory theory of justification.

The first burden is as follows. We want the best overall explanation of epistemic justification. The best explanations are highly general. So, we prefer to give a single explanation that covers *all* epistemic justification, apriori and empirical. But, a standard counterexample to reliabilism seems to show that empirically justified beliefs need not be the product of any reliable process. It seems possible that a brain-in-a-vat, which is fed a stream of misleading perceptual experiences indiscernible from some nonenvatted person's, can form completely justified beliefs. So, whatever explains why her beliefs are justified, it seems to not be that they are the product of a reliable process, much less a process guaranteed to be reliable. So, if we pursue a meaning-based view that explains apriori justification via a guarantee of reliability, then it seems we will have to settle for a disjunctive explanatory theory: we'll have to settle for a theory that says one thing

explains empirical justification, while a completely different thing explains apriori justification. Thus, the first heavy burden on pursuing such a view is to say why, contrary to how it just now seemed, we're not pursuing an explanatory theory that's less than best.

The second burden has been pressed by Boghossian. He draws on the other standard counterexample to reliabilism: it seems that an unwittingly reliable clairvoyant would not be justified in believing the deliverances of her faculty of clairvoyance. Boghossian (2003) suggests there can be reliable inferences that are essential to understanding some meanings, but which are as wildly irresponsible as the clairvoyant's beliefs. To illustrate this, he claims that there could be a concept, the concept *flurg*, such that fully understanding *flurg* requires *immediately* inferring from the arithmetical axioms that some numbers are flurg, and then from that *immediately* inferring Fermat's Last Theorem. (Here we're *referring* to this concept without *using* it ourselves.) This pair of immediate inferences leading to Fermat's theorem obviously doesn't transmit justification. The burden on the position in question, then, is to say why the inferences don't transmit justification in the same way the inference from *p&q* to *p* does. The only apparent way of doing this is to deny that there is any genuine concept which can only be fully grasped by making that wild leap to Fermat's theorem. So, the effective burden is to say why there is no such concept. That burden is difficult to shed, because the proposed inferences for *flurg* are truth-preserving; indeed, as an arithmetical theorem, Fermat's theorem is an apriori proposition. So, *flurg* cannot be dismissed as easily as *tonk* can be. We could add new constraints to our general theory linking concepts and conceptual roles, but it's unclear how to avoid introducing ad hoc constraints (i.e. the constraints cannot be motivated by the fact that they handle *flurg*).

These burdens certainly aren't the last word on whether a meaning-based view can explain justification by appealing to a guarantee of reliability (as the presentation of them as burdens, rather than refutations, was designed to emphasize). But this much pressure motivates us now to examine a meaning-based route to explaining epistemic justification that purports to do better.

Meaning-based Explanations of Epistemic Responsibility

Blind-Yet-Blameless Inference

How can the meaning-based theorist explain how beliefs are epistemically justified without drawing on the connection between meaning and reliability? The starting point of Boghossian's explanation is as follows. Even if apriori ways of acquiring justification are reliable, the unwittingly reliable clairvoyant shows that even when beliefs are reliably formed something can go wrong that leaves them unjustified. Boghossian suggests that what went wrong in the clairvoyant's case is that, in forming her beliefs in the way she did, she was being *irresponsible*. Boghossian's guiding idea, then, is that if we can explain why a thinker is being responsible when forming her beliefs in certain ways rather than others, then we will have explained why those are ways of acquiring justification. And that does seem plausible: The targeted notion of epistemic justification does seem to be tied to a kind of responsibility, which we can call *epistemic responsibility*. At a very minimum, epistemic responsibility is one philosophically worthwhile notion to aim to explain.

A simplistic approach to understanding epistemically responsible belief formation would be to explicate it as belief formation that the believer knows or somehow recognizes to be reliable. This would be one understandably natural initial reaction to the case

of the clairvoyant. Intuitively, the clairvoyant shouldn't have formed her beliefs *blindly* ("blindly" meaning "without any clue as to whether they were formed reliably"). But, Boghossian firmly rejects that approach, and I earlier mentioned his Carrollian argument against it. Perhaps if the clairvoyant knew that her faculty of clairvoyance is reliable, she'd be a responsible believer. But, according to the Carrollian argument, what *explains why* reasoning by *Modus Ponens* is responsible is *not* that I know my inference is reliable.

(To be sure, if I become philosophically reflective and wonder whether reasoning by Modus Ponens is reliable, the only way I can responsibly continue reasoning by Modus Ponens is if I give my question a positive answer. If I give a negative answer, or even if I adopt an agnostic view, then I am responsible for suspending all reasoning by Modus Ponens. But, this doesn't mean that higher-order knowledge of the reliability of reasoning by Modus Ponens *explains* why the reasoning is epistemically responsible. And, of course, it also doesn't mean that the responsible person on the street is required to engage in reflective epistemology in the first place.

There are philosophically interesting problems concerning *how* the reflective thinker could positively settle this higher-order question of Modus Ponens's reliability. Limitations of space prevent me from getting into it here. The classic discussion is Dummett (1973), further developed in Dummett (1991). Dummett argues, and Boghossian and many others have agreed, that we can *use* reasoning by Modus Ponens to infer the conclusion *that* reasoning by Modus Ponens is reliable. This is a so-called rule-circular argument, familiar also from the soundness proofs used in metalogic. I oppose the legitimacy of these rule-circular arguments in Dogramaci (2010).)

In looking around for some better way of explaining epistemic responsibility than an appeal to higher-order knowledge of reliability, Boghossian's next move is to offer the following argument. (He will not ultimately accept the argument, but it leads us to a better one.) Suppose a disposition to draw inferences of a certain form is essential to understanding the meaning of some basic term (or, if you prefer, grasping some basic concept). That term (or concept) would need to be used just to state the reliability claim for the associated form of inference. Therefore, prior to having any disposition to draw those inferences, it is not possible to evaluate the reliability of those very inferential transitions. And now an ought-implies-can style of maneuver comes: when it is not possible to even have the thought that your inference is reliable, it cannot be irresponsible of you to draw that inference blindly. Thus they are blind-yet-blameless inferences.

This *prima facie* appealing argument was endorsed in Boghossian (2001), but he brought it up again in Boghossian (2003) only to decide that it cannot adequately explain apriority. The problem is that if the argument is any good, it applies to all inferences that are essential to understanding a meaning, but then it applies to the irresponsible inferences for understanding *flurg*. So, we need a better argument, one that won't exonerate *flurg*'s inferences.

In the last sections of Boghossian (2003), he sketches a path to improving upon the argument for blind-yet-blameless inferences. Here's my elaboration of the key ideas from those last sections.

The clairvoyant was irresponsible, and an appealing explanation is that she's irresponsible because she formed her beliefs blindly when she did not have to. If we're to generalize that explanation, then we need to find some way in which someone reasoning by Modus Ponens must proceed blindly, while someone who draws the inferences essential to understanding *flurg* does *not* have to form her beliefs blindly. Boghossian is going to offer an idea for what that way is.

To get to that idea, he reviews a plausible point that Russell, Ramsey, Carnap, and David Lewis all made about the conceptual roles essential for understanding theoretical terms (and concepts) of science. Any inferences essential to grasping these terms hedge on the existence of a property or entity expressed by the term. Various proposals about how the hedging happens differ in their details, but the basic idea is simple. One illustrative way it might happen is that all the inferences essential to understanding, say, "gravity," must take place under a supposition: first we suppose there is a property of gravity, and only then we say gravity is the property with such-and-such a role. That role might even be the role of being the thing that makes such-and-such a set inferences truth-preserving (again, only supposing there is such a thing). If the conceptual roles are hedged in this way, both sides in a debate over the very *existence* of gravity may fully understand the terms of the debate. Hedging thus seems to be something responsible thinkers must do, at least wherever it is possible.

Boghossian doesn't explicitly put it this way, but we can even say that when inferences are hedged in the right way, they no longer count as blind; rather than being made "without any clue" as to whether they are reliable**,** they are made under a supposition that they are reliable. Seen in that way, hedging provides the key to understanding why someone who fully understands *flurg* is being irresponsible in just the same way the clairvoyant was. We wanted to say the clairvoyant formed beliefs blindly when she didn't have to. Now, to fully understand *flurg,* you do have to draw certain inferences, and you do have to draw them blindly. But what you don't *have to* do is fully understand *flurg* in the first place. *Flurg* has a sister concept, call it *flurg*,* that can serve any legitimate purpose *flurg* could serve but which requires only hedged inferences to be fully understood: first we suppose there exists a property, flurg*, which some numbers have and which makes the transition from axioms to Fermat's theorem valid; then, given the axioms, we immediately infer that some numbers are flurg*, and from that in turn we immediately infer Fermat's theorem. Those inferences, all made under a suitable supposition, aren't blind and certainly aren't irresponsible. Thus, the inferences associated with *flurg* are indeed blind, but they are not blameless: the inferences associated with *flurg*,* on the contrary, are *not* blind, and *are* blameless.

Finally, Boghossian claims that there is a minimal stock of concepts which a responsible thinker must be permitted to fully understand without engaging in any hedging. These are the concepts needed before you can hedge anything in the first place. He doesn't say exactly what's in this minimal stock, but he says the concept of the conditional certainly is, probably along with a few more very basic logical concepts. Fully understanding the meaning we actually express using "if" cannot require making any suppositions, since that would require already understanding a conditional thought. Rather, understanding the concept of the conditional requires a disposition to engage in certain reasoning flat out, including reasoning by Modus Ponens. There is no sister concept for the conditional; unhedged inferences are the only option when it comes to these conceptual abilities. Therefore, Boghossian concludes, since the only way to have a concept like this is to be disposed to draw certain inferences blindly, one is not thereby being irresponsible. These, and only these, are the blind-yet-blameless inferences.

Other Recent Views

In a, moment I'll turn to an objection to the above line of argument. First, I want to briefly mention recent work by other authors. The objection I will be raising applies to all these views.

Schechter and Enoch (2006) is primarily a criticism of meaning-based views. However, what Schechter and Enoch take issue with is only the thesis that the nature of epistemic justification can be *fundamentally* explained by any considerations concerning meaning. They take meaning-based theorists like Peacocke and Boghossian to task for false advertising about the role of meaning in the explanation of justification. Meaning-based theorists say that meaning considerations explain why we have justification, but then Peacocke appeals to reliability considerations, while Boghossian tacitly appeals to something like an ought-implies-can principle to supply the fundamental explanation; meaning only comes into the picture to account for why this or that is (guaranteed to be) reliable or (im)possible.

Once Schechter and Enoch turn to their own positive view, which they call *the pragmatic account* they also allow meaning to play that sort of ancillary role. Their view is that what fundamentally explains how our beliefs acquire justification is that we have an *epistemic obligation* to engage in certain cognitive projects, and this justifies the belief-forming methods that must be used if those projects are to succeed. For example, there might be an epistemic obligation to explain the world around you, and this, on their view, is what justifies a rule for, say, inferring the best explanation. They roughly characterize our basic epistemic obligations as obligations to explain, understand, inquire, reason, and deliberate about the world. Meaning comes into the picture because such obligations can only be successfully discharged if we are able to have conditional thoughts, and given CRS, this in turn requires us to draw certain inferences.

(Wright (2004) independently developed an extremely similar pragmatic view at the same time. I'm focusing on Schechter and Enoch (2006) here because Wright (2004) doesn't, at least not explicitly, discuss appealing to CRS.)

Schechter and Enoch even suggest that Boghossian is tacitly presupposing something like their view. They say:

> Indeed, it is plausible that Boghossian's discussion presupposes a version of the pragmatic account. . . . Boghossian claims that . . . [b]elief-forming methods that constitute unconditionalizeable [unhedgeable] concepts are justified . . . because such concepts are needed to conditionalize [hedge], and hence are needed to engage in inquiry. This suggests that on Boghossian's view the justification in all of these cases ultimately springs from the needs of rational inquiry, presumably a rationally required project. (708–9)

Boghossian never explicitly said certain methods are needed *to engage in inquiry*, only that they are needed to start hedging in the way a responsible thinker does. But Schechter and Enoch's suggestion seems a reasonable and fair way of elaborating the complete argument at work.

A line of argument again in the same family is endorsed in Wedgwood (2011). Wedgwood says:

> My current proposal, then, is that what makes the *primitively* rational cognitive processes *a priori* is that they are precisely those processes that we must have some disposition to engage in if we are to possess these basic cognitive capacities [where those capacities include "one's possession of each of the various concepts that one possesses"].

Here, "primitively rational" means justified, and not in virtue of the thinker's establishing their reliability using other arguments; "*a priori*" of course indicates that the justification is independent of experience. Having said *which* beliefs and inferences involve ways of acquiring apriori justification, Wedgwood then goes on to answer the attendant *why*-question in a way that's by now familiar to us:

> Can anything more be said to make it plausible that it is precisely these capacities that explain why the primitively rational processes are *a priori?* It may be relevant that these capacities seem to be necessary for even *considering* the questions that one actually asks oneself. Clearly, one's possession of the concepts that appear in these questions is a capacity that one must have in order to consider these questions. [All quotes from section 4 of manuscript.]

Wedgwood's overall view, which we're only scratching the surface of, is of course unique in a number of subtle ways; the same goes for Schechter and Enoch's positive view. But it's the common core elements that allow us to usefully group them together with Boghossian's account of blind-yet-blameless inference. And, as I'll now show, there is a single objection that threatens to bring down all these explanations of epistemic justification. (The objection threatens Wright (2004) too, but I cannot argue that here.)

(For a different set of complaints about meaning-based explanations of epistemic responsibility and justification, see Horwich (2005): chapter 6; and Horwich (2008). Since space is limited, I will not review Horwich's points here. I will rather develop an objection that has not been pressed against meaning-based views, though a paragraph in Williamson (2003): 252 anticipates the basic problem I develop.)

A Gap in the Meaning-Based Explanation of Responsibility

There's a familiar distinction, associated with Chomsky (see, e.g., Chomsky (1965)) between competence and performance. Most versions of CRS only say that one must be *competent* at drawing certain conclusions from certain premises in order to understand our ordinary logical concepts. What competence involves exactly is debatable. Maybe it just requires a simple disposition, maybe something more intellectually demanding, like rule-governed thinking, or maybe it's something else entirely. Whatever it requires exactly, on any plausible version of CRS, *performance errors* are perfectly consistent with understanding. After all, it's surely *possible* to affirm the consequent, but doing that requires understanding a conditional. I can still perfectly understand a conditional thought, *if p then q*, even if, say, on some occasion I irresponsibly affirm the consequent, inferring the conclusion *p* from beliefs in *q* and in *if p then q*. So, whatever level of understanding ordinary people have, it has to be consistent with such errors.

Now, the epistemologist's task is to explain facts about justification. The things that are justified or unjustified are beliefs, and the things that succeed or fail to transmit justification to beliefs are inferences. All this is on the performance side of the Chomskyan distinction. A cause for worry, now, is that the most that any of the arguments we heard shows is that if you're doing anything that's necessary for understanding basic logical thoughts, then you're being responsible. (You might even be meeting an obligation, as Schechter and Enoch have it.) But then this means what we heard could only explain

why it's responsible (or obligatory) to have *competence* with certain forms of inference. How do the arguments we've heard explain what we originally wanted to explain, the status of performances (their status as responsible, obligatory, justification transmitting, or whatever)?

The worry is not about having competence without ever exercising it. As mentioned early on, perhaps there could be a perfectly rational thinker who never believes *p* and *if p then q,* and so never has the opportunity to reason by Modus Ponens. Rather, the worry is about how the fact that an inference *manifests a responsible competence* should explain why the inference, the *performance,* is *itself* responsible (or obligatory, or justification transmitting, or whatever).

Let me now make the problem more concrete. Suppose, in a moment of weakness, I affirm the consequent. This is an irresponsible inference. If I *had,* in that instance, reasoned by Modus Ponens, then I *would have* made a responsible inference. Now, since affirming the consequent requires that I understand conditional thoughts, I must actually have whatever competence is required for understanding such thoughts. I've committed a performance error, but, I have also done everything necessary for understanding conditionals. But, if that's the actual case, then *it would not have helped enable me to understand conditionals had I reasoned by Modus Ponens;* I already understand conditionals. Therefore, my reasoning in the counterfactual scenario falls outside the class of what the meaning-based theorists argued is responsible. The meaning-based theorists' arguments fail to explain why my (counterfactual) inference is responsible. Furthermore, in a a case where I *actually* reason by Modus Ponens but easily *could have* affirmed the consequent, the meaning-based theorists cannot explain why my actual inference is responsible.

The basic source of the problem is that even if *having a competence* is responsible (obligatory, whatever), that doesn't entail that any *manifestation of that competence* is responsible (obligatory, whatever). Is this claim too strong? Could the meaning-based theorist resist it, perhaps by accusing it of "proving too much"? We might imagine the meaning-based theorist using the following analogy to argue that an explanation of why having a competence is good surely can always extend to explain why any competent performances are good. It's good to have a dust-free home. A dust-free home doesn't require any particular token dusting; it only requires a disposition or a habit of dusting occasionally. But, surely the goodness of a dust-free home explains why it was good that you dusted yesterday afternoon!

But, the analogy isn't apt. Every token dusting makes your home less dusty. Not every token instance of reasoning by Modus Ponens makes you more competent with the concept of the conditional.

Here's a better analogy for the relation between inferences and conceptual competence. Suppose animal overpopulation explains why it's good to issue a hunting license to someone. That explanation will entail that token killings are not all unjustified. However, it will not entail that all killings are justified. Once the population is thinned, there's no explanation left for why the next token killing is good; it's not good. In just the same way, if my competence at inferring *q* from *p* and *if p then q* is secured, then there's no explanation left for why any additional, "extra" inferences are responsible. There's a gap between the explanation of why the competence is good and the explanation of why *all* competent performances are good.

Meaning-based views have sparked a rich and valuable debate that spans nearly all of contemporary philosophy's main branches. As I've aimed to show here, their

viability depends on addressing some serious doubts about the specific proposals so far made, doubts concerning meaning's connections both to reliability and to justification. There's no doubt, however, that meaning-based theorists will continue to offer creative arguments for assigning meaning an important role in the theory of the apriori.

Related Topics

2.4 Concepts
2.5 Analytic Truth
6.1 Philosophy of Language for Epistemology.

References

Bealer, George (2000). "A Theory of the A Priori." *Pacific Philosophical Quarterly* 81(4): 1–30.

Belnap, Nuel (1962). "Tonk, Plonk and Plink." *Analysis* 22(6): 130–4.

Benacerraf, Paul (1973). "Mathematical Truth." *The Journal of Philosophy* 70(19): 661–79.

Bengson, John (2010). The Intellectual Given. Ph.D. thesis, The University of Texas at Austin.

Boghossian, Paul (1989). "The Rule-Following Considerations." *Mind* 98(392): 507–549.

—— (1996). "Analyticity Reconsidered." *Noûs* 30(3): 369–392.

—— (2001). "How Are Objective Epiatemic Reasons Possible?" *Philosophical Studies* 106(1–2): 1–40.

—— (2003). "Blind Reasoning." *Proceedings of the Aristotelian Society, Supplementary Volume* 77(1): 225–248.

BonJour, Laurence (1998). *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Carroll, Lewis (1895). "What the Tortoise Said to Achilles." *Mind* 4(14): 278–280.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chudnoff, Elijah (forthcoming). "The Nature of Intuitive Justification." *Philosophical Studies*.

Devitt, Michael (1998). "Naturalism and the A Priori." *Philosophical Studies* 92(1–2): 45–65.

Dogramaci, Sinan (2010). "Knowledge of Validity." *Noûs* 44(3): 403–432.

Dummett, Michael (1973). "The Justification of Deduction." *Proceedings of the British Academy* LIX: 201–232.

—— (1991). *The Logical Basis of Metaphyics*. Cambridge, MA: Harvard University Press.

Field, Hartry (1989). *Realism, Mathematics and Modality*. Oxford: Blackwell.

Hale, Bob and Crispin Wright (2001). *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Oxford University Press.

Harman, Gilbert (2003). "The Future of the A Priori." In *Philosophy in America at the Turn of the Century*, APA Centennial Supplement to the Journal of Philosophical Research, 23–34. Charlottesville, VA: Philosophy Documentation Center.

Horwich, Paul (2005). *Reflections on Meaning*. Oxford: Oxford University Press.

—— (2008). "Ungrounded Reason." *The Journal of Philosophy* 105(9): 453–471.

Huemer, Michael (2007). "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74(1): 30–55.

Kant, Immanuel (1781/1998). *Critique, of Pwe Reason*. Ed. Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.

Peacocke, Christopher (1993). "How Are A Priori Truths Possible?" *European Journal of Philosophy* 1(2): 175–199.

—— (1998). "Implicit Conceptions, Understanding and Rationality." *Philosophical Issues* 9: 43–88.

—— (2000). "Explaining the A Priori: The Programme of Moderate Rationalism." In Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the A Priori*, 255–285. Oxford: Oxford University Press.

—— (2004). *The Realm of Reason*. Oxford: Oxford University Press.

Prior, Arthur (1960). "The Runabout Inference-Ticket." *Analysis* 21(2): 38–9.

Quine, Willard Van Orman (1951). "Two Dogmas of Empiricism." *The Philosophical Review* 60(1): 20–43.

Schechter, Joshua and David Enoch (2006). "Meaning and Justification: The Case of Modus Ponens." *Noûs* 40(4): 687–715.

Sosa, Ernest (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.

Stalnaker, Robert (2003). *Ways a World Might Be*. Oxford: Oxford University Press.
Wedgwood, Ralph (2011). "Primitively Rational Belief-Forming Practices." In Reisner and Steglich-Petersen (eds.), *Reasons for Belief*, chapter 9. Cambridge: Cambridge University Press.
Weinberg, Jonathan, Shaun Nichols and Stephen Stich (2001). "Normativity and Epistemic Intuitions." *Philosophical Topics* 29(1&2): 429–460.
Williamson, Timothy (2003). "Understanding and Inference." *Aristotelian Society Supplementary Volume* 77(1): 249–293.
—— (2008). *The Philosophy of Philosophy*. Oxford: Blackwell.
Wright, Crispin (2001). "On Basic Logical Knowledge." *Philosophical Studies* 106(1–2): 41–85.
—— (2004). "Warrant for Nothing (and Foundations for Free)?" *Aristotelian Society Supplementary Volume* 78(1): 167–212.

6.6
NECESSITY AND MEANING

Gillian Russell

We can approach the topic of necessity and meaning by distinguishing three kinds of view about the relationship between modality—taken to include modal properties like necessity and possibility, and modal objects such as possible worlds and possibilia—and meaning. The first holds that modality is are rather mysterious and the best way to explain it (both metaphysically and epistemically) is in terms of more fundamental and accessible meaning properties. A famous example of this kind of view is the Linguistic Doctrine of Necessary Truth, according to which all necessary truth is to be explained in terms of analyticity (Ayer, 1936; Hempel, 1945; Goodman, 1955; Carnap, 1958).

A different kind of view holds instead that it is meanings which are mysterious, and the best way to understand them is in terms of more fundamental, and better understood, modal notions, such as those of worlds and possible objects. This approach reverses the direction of explanation and metaphysical dependence suggested by the positivists and a version can be found in Lewis (1976).

The third kind of view holds that semantic and modal notions do not mesh well enough for one to be entirely dependent on, or explained in terms of, the other. This view is of more recent provenance and reasons for it can be found in Thomason (1974), Kaplan (1989b), Soames (2004), King (2005), and Russell (2008).

There is room for more skeptical approaches than the three listed above; Quine had little time for full-blooded modal *or* semantic properties (Quine, 1951, 1953, 1966). But if there is no such thing as one, then it is straightforward to see that there will be no accounting for the other in terms of it, and so the focus of this chapter will be on what we are to make of the relationship between meaning and modality, on the assumption that both exist.

1

To say that something is necessary is to say that it could not have been otherwise and it is customary to distinguish two flavours. *De re* necessity is attributed to objects, as in:

(1) The number 2 is necessarily half of 4.
(2) Emeralds are necessarily green.
(3) There is something that has the property of being necessarily made of ice.

De dicto necessity, by contrast, is attributed to truth-bearers, such as statements, propositions or beliefs. Examples of ascriptions of de dicto necessity include:

(4) The sentence 'triangles have three sides' is necessarily true.
(5) It is necessary that 2+2=4.

Propositions are often said to be the *primary bearers* of de dicto necessity, where this is to say that if some other truth-bearer is de dicto necessary, then that is because it expresses a proposition which is. One reason for this is that propositions are often taken to be the primary bearers of truth, and from this it seems like a small step to say that they are the primary bearers of necessary truth. (Cartwright, 1962) Another is that it does not seem to be true that the sentence *2+2=4* is necessarily true; the sentence could have meant something else, in which case it might have been false. We might insist on only attributing necessity to the sentence *2+2=4* under the circumstances in which it expresses the proposition that 2+2=4, but then we may as well just say that it is propositions that are the primary bearers, and that at best sentences inherit the property when they express a necessarily true proposition.

De re necessity has sometimes been regarded as more metaphysically extravagant than the de dicto kind, but both varieties can be puzzling. To say *it is necessary that snow is white* seems to be to say no more about the way the world is than to say *snow is white*—it is not as if we are reporting on an empirically detectable halo of necessity which attaches to the snow's whiteness if it has it necessarily, and which is absent if it does not. One view is that when one attributes necessity, one is not saying something more about the way the world is, but rather saying something about the ways it could be. But what are these, if they are not facts about the way the world is? Thus there is a metaphysical puzzle.

There is also an epistemic puzzle. I think I know how one could learn that emeralds are green; one can *look at them*. But suppose it is true that emeralds aren't just green, but necessarily green. How would one learn that?

2 Linguistic Accounts of Necessity

An old idea is that both puzzles can be solved by appealing to meaning; necessity can be metaphysically grounded in the meanings of our expressions, and our knowledge of it is based on our the knowledge that speakers have of the meanings of those expressions. It is easier to see how this works in the case of de dicto necessity. First, we distinguish between two different kinds of sentence. *Synthetic* sentences true in virtue of what they mean and the way the world is, e.g. the sentence *snow is white* is true in part because of what it means and in part because snow has the colour it does. *Analytic* sentences, by contrast, are true in virtue of meaning alone. 'All triangles have three angles', to take an oft used example, will be true regardless of the way the world is.

The simplest version of the linguistic account says that a truth-bearer is necessary if and only if it is analytic. De re necessity, if we are to have any, is to be reduced to the de dicto kind. This is difficult, but at a first approximation we might suggest that if a sentence attributing a predicate *Fx* of an object a is analytic, then a has the property expressed by *F x* necessarily.

The linguistic approach is a tempting one. It is easy to understand in its basic outline and has the virtue of parsimony; it has no need of possible worlds, and we are not required

to possess a mysterious faculty of intuition for apprehending modal truths. I think it is fair to say, however that it is no longer the dominant view among philosophers of language. In the next section we will examine some of the reasons the view fell out of favour.

Quine's Critique

Quine was an influential critic of the linguistic account. Sometimes a complicated argument is made more succinct, and thus more tractable, if it is presented using symbols. So let's use '□' (pronounced like 'necessarily') as what Quine called a sentence operator; that is, like negation, '□' is an iterable operator which may be applied to either open or closed sentences. The following are then all well-formed:

(6) a) $\neg Fa$
    b) $\Box Fa$
(7) a) $\neg Fx$
    b) $\Box Fx$
(8) a) $\exists x \neg Fx$
    b) $\exists x \Box Fx$
(9) a) $\neg\neg\neg\exists x Fx$
    b) $\Box\Box\Box\exists x Fx$

In the first pair of sentences, the relevant operator is attached to an atomic sentence. In the second, the operator is instead attached to an open atomic formula, i.e. one containing at least one free variable. In (8) it appears within the scope of a non-vacuous quantifier, and in the last pair the operator is iterated, so that we have instances of ¬ and □ that fall within the scope of another ¬ or □, respectively.

About each construction we can ask: what does the construction mean? And what are its truth-conditions? If our constructions make sense, these questions should be answerable. Some argue that the second is just a more formal way of asking the first—i.e. that to give a sentence's truth-conditions is already to give its meaning—but intuitively, sentences with different meanings may have the same truth-conditions (e.g., *2+2=5* and *Triangles have four angles*) so even after we have answered the truth-condition question, we may not yet have an answer to the meaning question. Often an answer to the meaning question, however, will quickly give us an answer to the truth-condition question. For example, if $\Box Fa$ means that *Fa* is analytic, then we know that the sentence is true just in case *Fa* is analytic. Thus it looks as if the linguistic account of necessity can answer both questions for the sentences in (6). The problems start when we examine the more complicated constructions in (7)–(9).

Open Sentences

Applying the linguistic account flat-footedly to $\Box Fx$ suggests that it means that the open sentence *Fx* is analytic. But as normally understood, analyticity entails *truth* and open sentences like *Fx* are not objects of truth or falsity, except relative to particular assignments of objects to the variable *x*. Perhaps we could say that $\Box Fx$, like $\neg Fx$, is only true relative to an assignment, i.e. we could say that the open sentence was true *of* some objects and false *of* others. We know what is required in order for $\neg Fx$ to be true of some object; it must be that *Fx* is not true of it. What is required for $\Box Fx$ to be true of

an object? The natural answer is that the object must satisfy Fx necessarily, but if that is to be the right answer, then we have to make sense of de re necessity *before* we can make sense of sentences like (7b).

A linguistic account of de re necessity will say that objects inherit their modal properties (such as being necessarily *F*, or merely contingently G) from the analyticity of sentences. For example, one might say that $\Box Fx$ is true of an object o just in case $\Box Fa$ is analytic and *a* refers to o. This version of the approach exploits the analyticity of sentences containing singular terms which refer to the object. An alternative is to use predicates under which the object falls, i.e. one might say that $\Box Fx$ is true of o iff o is G and $\forall x(Fx \leftrightarrow Gx)$ is analytic.

The problem with both approaches is that objects may be referred to by many singular terms, and fall under many predicates, where some of these expressions form analytic sentences when substituted in the appropriate places and some do not. To use Quine's example, if we think that the sentences

b. (10) All mathematicians are rational.
(11) All cyclists are bipedal.

are both analytic, whereas the sentences

c. (12) All cyclists are rational.
d. (13) All mathematicians are bipedal.

are synthetic, we might use our strategy to conclude of any particular mathematician that she is necessarily rational but merely contingently bipedal, and of a particular cyclist that she is necessarily bipedal, but merely contingently rational. But the problem is that some cyclists are also mathematicians, and in such cases our strategy gives contradictory results: any cyclist mathematician is both necessarily rational and merely contingently rational.

The singular term strategy runs into similar problems. If $\Box Fx$ is true of an object o if and only if *t* refers to *o* and $\Box Ft$ is analytic, then if (within some particular context) we think that:

(14) The red book is red.
(15) The big book is big.

are both analytic, but

(16) The red book is big.
(17) The big book is red.

are merely synthetic, then our strategy tells us that the referent of 'the red book' is necessarily red, but contingently large, whereas the referent of 'the big book' is necessarily big, but contingently red. As before, it may turn out that the red book *is* the big book, and then our strategy leads to contradiction.

Several modifications might be attempted, but the most obvious ones turn out not to work: we might say that o is necessarily *F* iff *at least one* predicate G applying to it was such that $(Fx \leftrightarrow Gx)$ is analytic but a moment's thought shows that this leads to a trivialisation of de re necessity, making all objects possess all their properties necessarily. We might say that *o* is necessarily *F* iff *every single* predicate Φ applying to it was such that $\Box(Fx \leftrightarrow \Phi x)$ is analytic but this leads to *no* properties being possessed necessarily. One might say that only certain canonical kinds of predicates or terms can be

used for determining an object's de re modal properties, where we somehow ensure that any two canonical expressions delivered consistent results, but how are such canonical expressions to be identified, *unless* we say that a canonical predicate (say) is one which picks out a property which holds of the object necessarily—a strategy which amounts to abandoning the linguistic approach altogether (Burgess, 1997).

Non-Vacuously Quantified Sentences

Suppose we were to simply accept the problem above. One might think that this is only a minor concession for the linguistic account. Open sentences do not have truth-values anyway, and truth on an assignment is a technical notion from logic—it is not such a huge concession to have to admit that $\Box Fx$ does not have a truth-value relative to an assignment. But the costs are greater than this, because we have to be able to interpret open sentences like $\Box Fx$ in order to make sense of closed sentences in which they are embedded, such as $\exists x \Box Fx$. Standard truth-condition clauses in first-order logic tell us that for any formula Φx:

$\ulcorner \exists x \Phi x \urcorner$ is true iff $\ulcorner \Phi x \urcorner$ is true on some assignment.
$\ulcorner \forall x \Phi x \urcorner$ is true iff $\ulcorner \Phi x \urcorner$ is true on all assignments.

So if we are interested in making sense of a quantified sentence like $\exists x \Box Fx$ or $\forall x \Box Fx$ it tells us that:

$\exists x \Box Fx$ is true iff $\Box Fx$ is true on some assignment.
$\forall x \Box Fx$ is true iff $\Box Fx$ is true on all assignments.

So if we cannot make sense of an open-sentence being analytic relative to an assignment, then we cannot make sense of closed sentences in which the modal operator comes within the scope of a non–vacuously-binding quantifier either.

Iterated Modal Operators

A third problem arises for sentences which include multiple modal operators, as in (9b) above. We perhaps don't use these much in ordinary talk, but the ability to distinguish between such formulae and their consequences is one of the distinctive advantages of modal logic. If we had little need to consider iterated modalities, both quantified *and* sentential modal logic would seem undermotivated. What does $\Box\Box p$ mean on the linguistic account? $\Box p$ means that p is analytic. That suggests that $\Box\Box p$ expresses a metalinguistic claim:

(18) 'p is analytic' is analytic.

The problem is that on this interpretation it is hard to see why iterated modalities are of much interest. I am not sure whether

(19) 'Bachelors are unmarried' is analytic.

is analytic or not, but if it is, presumably all claims of the same form as (18) are true when an analytic sentence is substituted for p, and false where a synthetic sentence is. And if it

is not, then presumably all claims of the form of (18) are false. Either way, the truth-value of $\Box\Box p$ would be entirely determined by the truth-value of $\Box p$ and it looks as if all the fancier modal systems that allow for non-trivial iterated modalities are of little interest.

The Necessary A Posteriori

A more recent problem for the linguistic doctrine stems from the phenomenon of the necessary a posteriori (Kripke, 1980; Sidelle, 1989). At the height of logical positivism it was common for writers to collapse the ideas of necessity, a priority, analyticity, and linguistic truth (Burgess, 1997). But Kripke (1980) distinguished necessity (the metaphysical property of being such that it couldn't be otherwise) from a priority (the epistemic property of being knowable independently of experience) and argued that a truth could be necessary without being a priori. Kripke's most famous examples include:

(20) Hesperus is Phosphorus.
(21) Water is H_20.

although sometimes the more distracting questions (such as whether H_2O really expresses an a priori truth, and whether the semantic complexity of the expression 'H_2O' makes a difference) are avoided by using these related sentences:

(22) Mars is not Phosphorus.
(23) Gorse is furze. (Soames, 2001)

The most straightforward way to see what these examples are supposed to show is to start with Kripke's view that names are rigid designators, which is to say that they refer to the same object with respect to every possible world; if *Hesperus* refers to the planet Venus with respect to every possible world, and *Phosphorus* also refers to Venus with respect to every possible world, then, with respect to every possible world the proposition expressed by *Hesperus is Phosphorus* is simply that Venus is Venus—a necessary truth. Yet one might not expect to to be able to know this a priori. One would have to read it in a book, or make some astronomical observations. Hence it is an example of the necessary a posteriori. (Kripke doesn't put a lot of emphasis on distinguishing the objects of these properties in the original text, but one place where it was later done carefully and explicitly is (Soames, 2001).)

Such a result is also endorsed by standard modal logics, in which $(a = b) \supset (\Box(a = b))$ is a theorem, as is established by the following proof:

$(a = b) \supset (Fa \supset Fb)$	(Leibniz' Law)
$(a = b) \supset (\Box(a = a) \supset \Box(a = b))$	(Instance of Leibniz' Law)
$\Box(a = a)$	(necessity of identity)
$(a = b) \supset (\Box(a = b))$	

If names in English are suitable substituends for *a* and *b*, then it would seem that $\Box$(Hesperus is Phosphorus) is a logical consequence of *Hesperus is Phosphorus*:

A1 *Hesperus* = *Phosphorus* $\supset \Box$(*Hesperus* = *Phosphorus*)
A2 *Hespersus* = *Phosphorus*
A3 $\Box$(*Hesperus* = *Phosphorus*)

Since sentences which express a posteriori truths are normally thought to be synthetic, we seem to have an example of the necessary synthetic, which is prima facie inconsistent with the linguistic theory of necessity.

One attempt to preserve the theory in the face of this problem is (Sidelle, 1989). Sidelle suggests—plausibly enough—that the linguistic account need not strictly maintain that all necessary truths are analytic, so long as it maintains that the source of their necessity is analyticity. With regard to A3, for example, he suggests that though the *truth* of A3 depends on A2, the *modal force* of A3, as being necessary-if-true, comes entirely from the a priori, and analytic, A1. Though A3 is necessary and a posteriori, its necessity has a linguistic origin, in the analyticity of A1.

In a well-known review of Sidelle (1989), Yablo responds that Sidelle has underestimated the problem of explaining how the analyticity of A1 could explain its *own* necessity, far less that of A3. For, says Yablo, all the meaning of sentence can do is make it the case that it expresses a certain proposition. Whether or not that proposition is then true, or necessary, does not depend on the meaning of the sentence expressing it. Hence meaning alone is never sufficient to establish necessity. (Yablo, 1992) A line of response (though not one directed at vindicating the linguistic doctrine of necessary truth) can be found in Russell (2008).

The Contingent Analytic

A further problem for the linguistic account can be found by thinking about context-sensitive expressions like indexicals. An indexical is an expression that contributes different things to the proposition expressed by a sentence containing it in different contexts. Examples include *I*, *now*, *today*, *tomorrow* and *here*. If the sentence *It is hot today* is uttered on 21st June, it expresses a different proposition (says something different, with different truth-conditions) than it would have if it were uttered on 22nd June. Once we notice that sentences containing indexicals can be used to assert different propositions in different contexts, we have made room for the idea of a sentence with a special property, that of expressing different true propositions in different contexts, but nonetheless a true proposition in every context. Kaplan's example is:

(24) I am here now.

The idea is that whoever is speaking, wherever, and whenever, the proposition that gets expressed by (24) is a true one. People have proposed some interesting counterexamples to this claim but for present purposes the important thing is not that (24) is really the best example, but that we see that indexicality makes space for there to be a sentence with such a property. So for argument's sake, let's just assume *I am here now* is such a sentence.

The proposition that (24) expresses will usually be a contingent one. (Kaplan, 1989a) I *am* here now, but that didn't have to be the case—I could have gone in to the university instead of coming to this cafe, and then I would not have been here now. But the sentence doesn't have to express that proposition, and in other contexts it might express a necessary one. Suppose that God is necessarily in heaven at all times. Then when he utters "I am here now" he expresses a necessary truth. This theistic example raises some additional questions, but it has the virtue of making the basic problem clear in a simple fashion: indexical sentences express different propositions in different contexts. The same sentence may express a contingent proposition in one context, and a necessary one in another. But then what

explains the necessity of the proposition in the second context? It cannot be the meaning, or analyticity, of the sentence, since these were not sufficient for necessity in the first context (Russell, 2010).

3 Possible Worlds Accounts of Meaning

Some people are suspicious of the linguistic doctrine of necessary truth because they are suspicious of analyticity (Quine, 1935, 1951, 1954; Russell, 2008). If you do not believe in analyticity, then you will not think that necessity can be accounted for by appeal to it and that will leave you with two options: find an alternative account, or be skeptical about necessity.

Skepticism about necessity is difficult to argue against, but it has become an increasingly uncomfortable position to hold. It looks likely that modality has an important role to play in science and philosophy, the development of modal logics has lent modality a veneer of respectability, and it is clear that talk of necessity and possibility is not limited to philosophical and scientific discussion but is prevalent in ordinary conversation as well. For these reasons skepticism about modality has significant costs.

Sometimes it is not a refutation of a skeptical view that tempts philosophers away from it, but rather the anticipated fruitfulness of approaches that begin from less skeptical assumptions. The work of Montague, Kamp, Lewis, Stalnaker and others has encouraged the idea that rather than giving an account of modality in terms of meaning, we might instead give an account of meaning in terms of modality (Lewis, 1976; Thomason, 1974; Stalnaker, 1999).

We might think of this as an enhancement of some of Frege's views (Frege, 1892, 1918). Frege held that expressions—including entire sentences—had referents. The referent of a singular term was an object, the referent of a whole sentence a truth-value. The referents of complex expressions are determined by the referents of their parts, so the referent of a unary predicate needs to be something that could be put together with an object to yield a truth-value, and a natural choice is a function from objects to truth-values. Along the same lines, first-order quantifiers are things that can be put together with a unary-predicate to form a sentence and so their referents will be functions from the referents of predicates to truth-values, that is functions, from (functions from objects to truth-values) to truth-values.

This is a powerful framework. We started with just two basic types of thing—objects and truth-values—as the extensions of singular terms and sentences, and then used the machinery of set theory to generate complex, non-basic types to serve as extensions for other kinds of expression.

But the idea runs into problems when we consider sentences that contain that-clauses, most famously, propositional attitude ascriptions such as:

(25) Mary believes that *snow is white*.
(26) Mary believes that *grass is green*.

Given what we have said so far, the contribution that the italicised parts make to the extensions of sentences (25) and (26) should be the same, since the extension of both *snow is white* and *grass is green* is the truth-value true, and given that the containing sentences are the same in other respects, (25) and (26) must have the same truth-value. Yet, as Frege realised, that is the wrong result: the truth-values of (25) and (26) could be different.

His solution was to say that expressions had two kinds of referent, an ordinary *direct* referent and an *indirect referent* which it would take in certain contexts, such as attitude ascriptions. Frege held that the indirect referent of an expression is its *sense*. Frege's senses were introduced for reasons independent of this problem (Dummett, 1981), and they play many roles—for example, the sense of a sentence is the thought (or proposition) it expresses (Salmon, 1982). But what is critical to the solution to the present problem is that *grass is green* and *snow is white* have different senses and hence different indirect referents, which allows the truth-values of (25) and (26) to come apart.

We could think of Frege's solution to the problem as involving the introduction of a new basic type—senses. So far, so good. What is a sense? (Or using less Fregean terminology, a propositional content?) For Frege they were denizens of the third realm, neither empirical objects, nor subjective ideas, but eternal abstracta (Frege, 1918). Not everyone has found this satisfactory. But even if one has no problem with it on grounds of parsimony, it is understandable that we should want to say more about such things if they are to be put to serious explanatory use (Davidson, 1967; Lewis, 1976).

Although Frege didn't pursue this line himself (and had little interest in necessity or possible worlds) a popular approach is to characterise content as a function from possible worlds to extensions. (It invites problems to use the multiply connoting *senses* for such things, so let's call them *intensions*.) What we will add to our list of basic types then is not unanalysed senses, but *possible worlds*. Possible worlds are ways the world could be, but what they are beyond that is contentious. For Lewis they were concrete universes, as real as our own. For others they are properties (Stalnaker, 1976) (Soames, 2001, 23) or fictions (Rosen, 1990) and there are even those who hold that for present purposes, figuring out what possible worlds are doesn't matter; Stalnaker characterises the decision to theorise using possible worlds as simply a decision to theorise at a certain level of abstraction (Stalnaker, 1999, 79).

Yet one could be forgiven for thinking that this is not really moving us in the right direction. It is true that meanings or senses are in need of a philosophical account, but it is not obviously helping to explain them in terms of possible worlds. What the meaning of the sentence *Theatetus flies* is might be a little obscure, but it's familiar enough an idea—certainly meaning is a part of the ordinary, everyday talk of the folk—and it is rather surprising to learn of such a familiar thing that it is in fact a function from possible worlds to extensions.

But to admit this is perhaps only to admit to having little native sympathy with the thesis that meanings are more mysterious that possibilities. Native sympathies should be relinquished where there is good reason, and the failure of the linguistic doctrine suggests that it is worth considering a rival approach and, in fact as we will see, the project of possible worlds semantics has seen considerable success.

Adding possible worlds to our list of basic types adds many non-basic types, such as functions from possible worlds to truth-values and functions from functions-from-possible-worlds-to-truth-values to truth-values. Descriptions like this are hard to parse and some formalism is useful for referring to functions that can take other functions as their arguments and values. We follow convention in using the lowercase letter *e* to stand for the basic type *objects* (you can think of it as *e* for *entity*) and lowercase *t* for the basic type *truth-values*. If σ and τ are both types, then $\langle \sigma, \tau \rangle$ is the type of functions from extensions of the type of σ to extensions of the type of τ. For example, $\langle e, t \rangle$ is the type of functions from objects to truth-values (intransitive verbs have extensions of this type) and $\langle \langle e, t, \rangle, t \rangle$ is the type of functions from [functions from objects to truth-values] to

truth-values—the type of quantifiers. The intension of a name will be of type $\langle w, e\rangle$, that of a one-place predicate is $\langle w, \langle e, t\rangle\rangle$ and that of a sentence $\langle w, t\rangle$.

We can now draw some useful distinctions that we were unable to draw before. Here are three examples:

Some pairs of non-synonymous predicates are co-extensional—they are satisfied by all the same objects. A famous example uses the words *renate* and *cordate* (Quine, 1986). Since all animals with kidneys have hearts, and vice versa, these two expressions correspond to the same function from objects to truth-values $\langle e, t\rangle$. But our possible worlds machinery allows us to tell them apart: the intension of *renate* is not its extension, but a function of the type $\langle w, \langle e, t\rangle\rangle$. Since there are possible worlds in which there are animals which are renates but not cordates, there are pairs of objects and possible worlds for which the intension of *renate* yields true, but the intension of *cordate* yields false. Hence these are not the same function—the two words have different meanings, as desired.

Another advantage of using possible worlds to characterise meanings is that they allow us to distinguish rigid from non-rigid designators. A rigid designator is an expression that has the same denotation with respect to every possible world (Kripke, 1980). This property does important work in explaining the characteristics of certain sentences and expressions (for example in explaining why identity statements between co-referring names express necessary truths) but it is an intensional property, and without reference to modal notions like possible worlds we cannot distinguish rigid from non-rigid designators: the rigid designators are exactly the ones whose intension is a *constant* function from possible worlds to objects.

A third advantage, of course, is that we can use intensions to distinguish the indirect referents of sentences embedded in propositional attitude ascriptions—even where the embedded sentences share a truth-value—and hence avoid the unwelcome conclusion that the truth-values of (25) and (26) must be the same if the truth-values of *snow is white* and *grass is green* are the same.

Problems for Possible Worlds Accounts

The benefits of possible world semantics are significant, but the simple approach outlined above faces problems that are strikingly similar to the ones it solves (Kaplan, 1989b), (Soames, 1998), (Soames, 2004) (King, 2005). Consider these correlates of (25) and (26):

(27) Mary believes that *triangles have three sides*.
(28) Mary believes that *there is no highest prime*.

Both embedded sentences express necessary truths and so the indirect referent of each—on the theory we are considering—is the constant function from each possible world to the true. Since every other expression in (27) and (28) is shared between the sentences, our theory holds that the two belief ascriptions share a truth-value. But this is wrong. Mary might believe that triangles have three sides, but not yet have studied sufficient arithmetic to have any belief about whether or not there is a highest prime.

There is also a problem with necessarily co-extensional predicates. All and only quadrangles are quadralaterals, and this is not merely contingently the case, but necessarily. Given this, the words *quadrangle* and *quadralateral* are predicted to have the same content: the function that maps each quadralateral in each possible world to the true

and everything else to the false. Intuitively, however, they express different properties: a quadrangle is a figure with four corners, whereas a quadralateral is a figure with four sides; it is a matter of geometrical fact, rather than a mere consequence of their meanings, that the two properties are necessarily co-extensive.

Finally, our theory is not fine-grained enough to distinguish directly referential expressions from mere rigid designators (Kaplan, 1989a; King, 2005; Russell, 2008). A directly referential expression is one whose extension is not determined by its content. Neo-Millian names provide one of the clearest examples. The referent of a neo-Millian name like *Mary* might be established via an initial baptism, and then passed on from speaker to speaker through conversation (the details are non-trivial but not critical here) but when a speaker utters a sentence like *Mary is tall*, what the name contributes to the proposition that gets expressed is not some descriptive information such as that expressed by 'the person baptised at such and such a date' or 'the firstborn of Mr and Mrs Green' but rather Mary herself (making the proposition expressed a Russellian one).

Such expressions are rigid designators. Since the referent of the name is already contained in the content, there is no opportunity for the content to select different referents from different possible worlds. By contrast, the description *the person baptised at such and such a time* has a descriptive content which will pick out different people given different possible worlds (Kaplan, 1989a). Descriptions can be rigidified, for example by using the operators *actual* or *dthat*. Yet intuitively the *contents* of *Mary* and *the actual person baptised at such and such a time* are different, since one is a person, and the other not. If we characterise content using functions from possible worlds to extensions, we leave something out. There are several ingenious approaches to making possible worlds accounts more fine-grained in order to respond problems like those above, including the use of:

1. situations (incomplete possible worlds) (Barwise and Perry, 1998)
2. impossible worlds (inconsistent possible worlds) (Takashi, 1988; Lewis, 1998)
3. two-dimensional semantics and epistemically possible worlds (Chalmers, 2004; Soames, 2004)
4. combination possible worlds/structured propositions approaches (Lewis, 1976) (King, 2005)

But three facts should make us cautious if all of these accounts characterise content as a function from possible worlds-type objects to extensions. The first is that it is hard to see how they will be able to distinguish rigidity from direct reference, since two terms may be associated with the *same* function from (appropriately conceived) possible worlds-type objects to extensions, whilst one does so via a function that is a part of its content, and another does so directly. The second is that the problems the simple accounts ran into were similar to the problems they were intended to solve. Such 'revenge' problems suggest that the attempted solution was never adequate in the first place; an adequate solution ought to solve these kinds of problems together. Finally, one may also suspect that some of the techniques that these solutions add to possible world semantics—such as the structuring of propositions—were sufficient to solve the problems on their own, without the addition of possible worlds.

4 The independence of meaning and modality

From here, the Vienna Circle's time seems like an austere one. Today's philosopher tends to be more liberal about his metaphysical commitments, and concurrently more dependable when it comes to drawing certain distinctions. The result has been a decline in the felt strength of the motivation for the linguistic doctrine of necessary truth, and an increase in the appreciation of the problems that the doctrine faces. But something of austerity of the positivists' spirit remains, this time in the reluctance to treat meanings as respectable in their own right, and the resulting push to account for them in modal terms. But whereas meanings often determine functions from possible worlds to extensions, the austere thesis that functions from possible worlds to extensions is *all there is* to meanings looks increasingly inadequate.

References

Ayer, A. J. (1990/1936). *Language, Truth and Logic*. Penguin Books, Harmondsworth.

Barwise, J. and Perry, J. (1998). *Situations and Attitudes*. CSLI, Stanford, CA.

Burgess, J. P. (1997). Quinus ab omni nævo vindicatus. *Canadian Journal of Philosophy, Supplementary Volume 23: Meaning and Reference*.

Carnap, R. (1958). *Empiricism, Semantics and Ontology*, pages 205–221. The University of Chicago Press, Chicago and London, 2nd edition.

Cartwright, R. (1962). Propositions. In Butler, R., editor, *Analytical Philosophy*, pages 81–103. Blackwell, Oxford.

Chalmers, D. J. (2004). The foundations of two-dimensional semantics. In Garcia-Caprintero, M. and Macia, J., editors, *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press, Oxford.

Davidson, D. (1985/1967). Truth and meaning. In Martinich, A. P., editor, *The Philosophy of Language*, pages 98–121. Oxford University Press, Oxford, 4th edition.

Dummett, M. (1981). *Frege: Philosophy of Language*. Harvard University Press, Cambridge, Massachusetts, second edition.

Follesdal, D. (1969). Quine on modality. In Davidson, D. and Hintikka, J., editors, *Words and Objections: Essays on the Work of W.V. Quine*, pages 175–185. Reidel, Dordrecht.

Frege, G. (1956/1918). The Thought. *Mind*, 65(259):289–311.

Frege, G. (1985/1892). On Sense and Nominatum. In Martinich, A. P., editor, *The Philosophy of Language*. Oxford University Press, Oxford, 4th edition.

Goodman, N. (1955). *Fact, Fiction and Forecast*. Harvard University Press, Cambridge, Mass.

Hempel, C. G. (1964/1945). On the nature of mathematical truth. In Benacerraf, P. and Putnam, H., editors, *Philosophy of Mathematics: Selected Readings*, pages 85–112. Prentice-Hall Inc, Englewood Cliffs, New Jersey.

Kaplan, D. (1989a). Afterthoughts. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes from Kaplan*. Oxford University Press, New York.

Kaplan, D. (1989b). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes from Kaplan*. Oxford University Press, New York.

King, J. C. (Summer 2005). Structured propositions. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University.

Kripke, S. A. (1980). *Naming and Necessity*. Blackwell, Oxford.

Lewis, D. (1976). General semantics. In Partee, B., editor, *Montague Grammar*, pages 1–50. Academic Press, New York.

Lewis, D. (1998). *Relevant Implication*, chapter 8, pages 111–124. Cambridge University Press, Cambridge.

Quine, W. V. O. (1935). Truth by convention. In *The Ways of Paradox and other essays*, chapter 9, pages 70–99. Random House, New York.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60:20–43.

Quine, W. V. O. (1953). *Reference and modality*, chapter 8. Harvard University Press, Cambridge, MA.

Quine, W. V. O. (1954). Carnap and logical truth. In *The Ways of Paradox and other essays*, chapter 10, pages 100–125. Random House, New York.

Quine, W. V. O. (1966). *Three grades of modal involvement*, chapter 15. Harvard University Press, Cambridge, MA.
Quine, W. V. O. (1986). *Philosophy of Logic*. Harvard University Press, Cambridge, Mass.
Rosen, G. (1990). Modal fictionalism. *Mind*, 99(395):327–354.
Russell, G. (2008). *Truth in Virtue of Meaning: A Defence of the Analytic/Synthetic Distinction*. Oxford University Press, Oxford.
Russell, G. (2010). A new problem for the lingusitic doctrine of necessary truth. In *New Waves in Truth*, pages 267–281. Palgrave MacMillan.
Salmon, N. (1982). *Reference and Essence*. Blackwell, Oxford.
Sidelle, A. (1989). *Necessity, Essence and Individuation: A Defence of Conventionalism*. Cornell University Press, Ithica and London.
Soames, S. (1998). Facts, truth conditions and the skeptical solution to the rule-following paradox. *Language, Mind and Ontology*, 12:313–348.
Soames, S. (2001). *Beyond Rigidity*, chapter 9-11. Oxford University Press, Oxford.
Soames, S. (2004). *Reference and Description: the case against two dimensionalism*. Princeton University Press, Oxford.
Stalnaker, R. (1999). *Context and Content*. Oxford University Press, Oxford.
Stalnaker, R. C. (1976). Possible worlds. *Nos*, 10(1):65–75.
Stein, H. (1992). Was Carnap entirely wrong, after all? *Synthese*, 93:275–295.
Takashi, Y. (1988). Beyond possible worlds. *Philosophical Studies*, 53:175–204.
Thomason, R. (1974). Introduction. In Thomason, R., editor, *Formal Philosophy: Selected Papers of Richard Montague*, pages 1–69. Yale University Press, New Haven.
Yablo, S. (1992). Review of Necessity, Essence, and Individuation: A defense of conventionalism. by Alan Sidelle. *The Philosophical Review*, 101(4):878–881.

6.7
PROPOSITIONAL ATTITUDE REPORTS

David Shier

Propositional attitudes are cognitive states such as believing, desiring, doubting, and hoping. Propositional attitude reports (or ascriptions)—i.e., sentences reporting the propositional attitudes of individuals—are central to our psychological discourse and to our understanding of the world, since in order to explain and predict behavior, we must appeal to information about the beliefs, desires, etc. of ourselves and others.

These attitude reports (often generically called "belief reports") are of special interest to philosophers of language because a number of particularly challenging problems arise in connection with these linguistic constructions, including (among others) issues about the substitution of co-referring terms into attitude reports, the rationality of belief and other attitude states, the status of so-called *de re* thoughts, and cognitive anti-individualism.

The sentences we use to report propositional attitudes come in various forms, including sentences such as the following (where 'believes' can be replaced by other cognitive attitude verbs).

1. Faith believes that koalas are herbivorous
2. Madeline believes Zorn's Lemma
3. Isabel believes Phyllis

While these are all attitude reports, the philosophical focus is primarily on locutions such as (1) that characterize the *contents* of the reported beliefs, especially when the report is constructed from a subject, a propositional attitude verb, and an embedded declarative sentence prefaced with the complementizer 'that'. Discussion focuses on such sentences both because they are so central to our explanations and predictions of behavior and because the aforementioned puzzles about attitude reports arise primarily with respect to reports of this sort.

The Propositional Theory

Since the pioneering work of Gottlob Frege (1892), the received account of propositional attitudes has been what might be called *The Propositional Theory* (its orthodoxy reflected in the very label "*propositional* attitude"). It comprises the following two claims.

- ***Relational Account of Attitudes*: Propositional attitudes are relations that individuals can bear to propositions.** (Propositions are also typically taken to be the *individuating* contents of beliefs, so that to share a belief is to believe the same proposition.)
- ***Relational Analysis of Attitude Reports*: A (*de dicto**) attitude report of the form *'S believes (desires, hopes, etc.) that P'* ascribes to S the relational property of bearing the relevant attitude (belief, etc.) to the proposition expressed by the embedded sentence of its 'that'-clause.** ('That'-clauses are also typically taken to be referring expressions designating the propositions expressed by their embedded sentences, thus making the logical form of the attitude report straightforwardly relational.)

  * See "De Re/De Dicto Distinctions" below.

The Propositional Theory provides an integrated, plausible and rather robust explanation of many of the facts about cognitive states and the sentences we report them with. Discourse about beliefs, etc. strongly suggests a commitment to abstract entities like propositions as the contents of these states. For example, Existential Generalization appears to apply validly to attitude reports, as in the inferences (4)-(5) and (6)-(8). And as Jerry Fodor (1978) says, "EG may not be *criterial* for ontological commitment, but surely it is a straw in the wind."

4. Phyllis believes that three is prime and four is composite
5. There is *something* that Phyllis believes
6. Bernard believes that virtue is its own reward
7. Sophie believes that virtue is its own reward
8. Therefore, there is something that Bernard and Sophie both believe

Furthermore, *what* Phyllis believes must have a number of important semantic properties such as being *true*, being *about* three and four, *entailing* that three is prime, etc.—the very sorts of semantic properties that propositions are standardly thought to have. The Relational Theory's appeal to propositions also straightforwardly explains the validity of various other common inferences, such as the following.

9. Daisy said that squirrels are omnivorous
10. Otis believes what Daisy said
11. Therefore, Otis believes that squirrels are omnivorous

Propositionalism is not without its challengers (e.g., the sententialism of Rudolf Carnap (1947), W.V.O. Quine (1956), and Donald Davidson (1968), as well as Michael McKinsey's (1986, 1999) property theory and Kent Bach's (1997) descriptivism). Furthermore, the nature of propositions is itself the subject of debate, for example as to whether they are best conceived as highly structured pieces of information or as sets of possible worlds, although the latter view faces significant difficulties, summarized nicely by Soames (1987: 197–200). Nevertheless, Propositionalism and its close variants dominate the literature.

Frege

Among the most challenging puzzles about cognitive attitude reports are those involving the substitution of co-referring terms into 'that'-clauses. Efforts to resolve these have played a prominent role in shaping theories of the attitudes as well as semantic theories of various parts of speech, such as names.

As Frege (1892) realized, if the truth value of a sentence is a function of what its constituent expressions refer to, then substituting one expression for another referring to the same item should not produce a sentence with a different truth value. (Jennifer Saul (2007), though, has argued that substitution puzzles can arise outside of attitude reports, even in simple sentences.)

For example, consider the Spider-Man comics (and treat them as if they were factual for present purposes). Since 'Spider-Man' and 'Peter Parker' refer to the same person, and (12) is true, then it seems (13) must be true as well.

12. Peter Parker lives in New York
13. Spider-Man lives in New York

But Frege also realized that certain linguistic contexts seem not to permit such substitution *salva veritatae* of co-refererring expressions. Prominent among these so-called *opaque* contexts are the 'that'-clauses of some propositional attitude reports. For example, Peter's boss is unaware of Spider-Man's secret identity, so even though (14) is true, there are strong reasons for thinking (15) must be false. (Among these reasons is the fact that Jameson himself would not assent to 'Peter Parker is a vigilante'.)

14. Jameson believes that Spider-Man is a vigilante
15. Jameson believes that Peter Parker is a vigilante

In addressing the puzzle he discovered about informativeness (e.g., how can a true statement of the form 'a=b' be informative when 'a=a' is not?), Frege famously concluded that there must be two dimensions to an expression's meaning –its *Bedeutung* or reference (the entity singled out) and its *Sinn* or sense (a way of singling out the referent). Frege's solution to the problem of substitution in attitude reports also makes use of his theory of *Sinn* and *Bedeutung* by postulating that inside the scope of a cognitive attitude verb, the reference of an expression shifts from its customary reference to its customary sense.

Thus, on the Fregean view, it is not problematic that (14) and (15) should diverge in truth value, because the embedded tokens of the names 'Spider-Man' and 'Peter Parker' are not co-referential and instead refer to two distinct senses (the web-slinging costume-wearing way of picking him out and the young photographer way). It is worth noting that on Frege's view, then, substitution of co-referentials *salva veritatae* technically only *appears* to fail, since the name tokens inside 'that'-clauses are not actually co-referential.

Frege's solution, while elegant, faces serious difficulties. Among the major problems specific to the Fregean analysis of attitude reports is one involving anaphoric pronouns, such as the occurrence of 'he' in (16).

16. Jameson believes that Peter Parker is cowardly, but he really isn't

On Frege's analysis of attitude reports, the token of 'Peter Parker' in (16) refers not to Peter himself, but to a sense—a way of picking out Peter. Since the anaphoric pronoun 'he' inherits its reference from the token of the name, the Fregean analysis would say that the pronoun also refers to this sense, which is plainly wrong, since it surely isn't the sense of 'Peter Parker' who is said not to be cowardly, but Peter himself.

Furthermore, the Fregean analysis violates the intuitively plausible principle of Semantic Innocence, according to which the semantic contribution made by a term is the same whether or not the term is embedded in a 'that'-clause.

Russell

Russell's (1905 and 1912) solution to substitution problems originates in his Theory of Descriptions and differs significantly from Frege's. While Russell, unlike Frege, held that logically proper names are genuinely singular terms (i.e., terms that are directly referential in the sense that they contribute just their referents to the propositions expressed by sentences containing them), he argued that other types of terms, including definite descriptions, could not be genuinely singular. One of the lines of thought that led him to this view had to do with substitution failure.

Consider attitude reports with descriptions in their 'that'-clauses. Even though the author of *Nobody's Fool* (Richard Russo) is the author of *Straight Man*, it is entirely possible for (17) and (18) to have different truth values.

17. Otis believes that the author of *Nobody's Fool* won the Pulitzer
18. Otis believes that the author of *Straight Man* won the Pulitzer

Russell's explanation of substitution failure, unlike Frege's, did not entail that terms in attitude reports function differently than elsewhere. Instead Russell inferred from substitution failure (and other phenomena) that sentences containing descriptions have unexpectedly complex logical forms. He famously argued that despite their surface grammatical appearances, sentences of the form *The F is G* do not have simple subject-predicate form and instead have underlying logical forms involving multiple quantifiers, with the following truth conditions.

19. There is at least one F, there is at most one F, and whatever is F is G

Thus, on Russell's Theory of Descriptions, the propositions expressed by the 'that'-clauses of (17) and (18) are given by the following—where 'Nx,' 'Sx' and 'Px' abbreviate 'x authored *Nobody's Fool*' 'x authored *Straight Man*' and 'x won the Pulitzer.'

20. $\exists x\,(Nx \,\&\, \forall y(Ny \rightarrow y=x) \,\&\, Px\,)$
21. $\exists x\,(Sx \,\&\, \forall y(Sy \rightarrow y=x) \,\&\, Px\,)$

Therefore (17) and (18) correspond to (22) and (23), where '<S>' indicates the proposition expressed by 'S'.

22. Believes (Otis, $<\exists x\,(Nx \,\&\, \forall y(Ny \rightarrow y=x) \,\&\, Px\,)>$)
23. Believes (Otis, $<\exists x\,(Sx \,\&\, \forall y(Sy \rightarrow y=x) \,\&\, Px\,)>$)

But (22) and (23) clearly report belief relations to entirely different propositions, so there is no reason to expect them to share truth values. Thus Russell neatly resolves the substitution problem, at least with regard to descriptions.

It should be noted that Russell held a view somewhat more complicated than the straightforward Relational Analysis. Russell (1912) stated that belief reports do not ascribe binary relations between believers and propositions, but rather report more complex relations among the believers and the constituents of the proposition believed. In Russell's example, Othello's belief that Desdemona loves Cassio is a relation among Othello, Desdemona, Cassio, and the relationship of loving. But this complication can be ignored for our present purposes.

It should also be noted that while (17) and (18) are interpreted above in the most natural way (with the descriptions understood as characterizing how Otis thinks of the described individual, and with the associated quantifiers, etc. occurring in small scope, or *secondary occurrence*, as Russell called it, relative to the belief operator), there are attitude reports—including less-natural readings of (17) and (18)—in which descriptions take large scope (or *primary occurrence*, as Russell called it) relative to belief operators. These reports would have logical forms such as those in (24) and (25), with respect to which the substitution of co-referring descriptions is uncontroversially valid. Sentence (24) states that the author of *Nobody's Fool* is such that Otis believes him to be a Pulitzer winner, and (25) states that the author of *Straight Man* is such that Otis believes him to be a Pulitzer winner. (For further discussion, see "*De re*/*de dicto* Distinctions" below.)

24. $\exists x\,[Nx \,\&\, \forall y(Ny \rightarrow y=x) \,\&\, \text{Believes (Otis, <}Px\text{>)}]$
25. $\exists x\,[Sx \,\&\, \forall y(Sy \rightarrow y=x) \,\&\, \text{Believes (Otis, <}Px\text{>)}]$

Let us now consider how Russell analyzed attitude reports with *proper names* occurring in small scope, inside the scope of the cognitive attitude verbs.

14. Jameson believes that Spider-Man is a menace to New York
15. Jameson believes that Peter Parker is a menace to New York

If names like 'Peter Parker' and 'Spider-Man' function as genuine singular terms, as Russell thought logically proper names do, he would have to regard (14) and (15) as expressing the same proposition and hence the same in truth value.

Russell's (1912) surprising solution was not to challenge the Relational Analysis or the claim that logically proper names are genuine singular terms. Rather, he postulated that ordinary proper names are simply not *logically* proper names and that they function instead as if they go proxy for definite descriptions. So the embedded sentences of (14) and (15) would express propositions like those expressed by (26) and (27).

26. The web-slinging, spider-costume-wearer is a menace to New York
27. The young photographer who works for the Bugle is a menace to New York

Therefore, just as with definite descriptions, the belief reports (14) and (15) would attribute to Jameson belief of very different propositions, thus explaining how their truth values can diverge.

Russell's solution (often labeled *The Description Theory of Names*) has the virtue of being well integrated into his broader philosophical views. It is a key part of his solutions to

several other important problems, such as Frege's puzzle about informativeness, problems about negative existence claims, and issues about non-referring names. Additionally, it is well motivated by his epistemological commitments, such as his Principle of Acquaintance which states that sentences we understand must express propositions composed only of items we can directly apprehend—i.e., items of our immediate acquaintance. According to Russell (1912), the only sorts of things with which we have this acquaintance are properties and relations, our own sense data, and (perhaps) ourselves. Therefore, a proper name of any other sort of item (such as an external physical object or another person) cannot contribute its referent to the proposition, for we could not understand the propositions expressed by such sentences, and clearly we do. But he held we can be directly acquainted with the properties and relations that descriptions contribute to the propositions.

Russell's solution is superior to Frege's in several respects. For instance, it does not violate Semantic Innocence, since on his view names and other terms inside 'that'-clauses make the same contributions to propositions that they make elsewhere. Neither does Russell's solution face the problems Frege's does in accounting for anaphoric pronouns.

Nonetheless, there are persuasive reasons for rejecting Russell's Description Theory of Names. Principal among these are arguments advanced by Saul Kripke (1972), Keith Donnellan (1970), and others against the description theory of names and in support of the Direct Reference semantic theory that supplanted it. The two main approaches to propositional attitude reports in the wake of the Direct Reference theory are Naïve Russellianism and The Hidden-Indexical View.

Naïve Russellianism

Naive-Russellianism—so-called because it resembles Russell's early semantic views—holds that ordinary proper names and indexicals normally function as Russell thought *logically* proper names did—i.e., as genuine singular terms, contributing only their referents to the expressed propositions, even when these terms designate external contingent objects.

Naive-Russellianism furthermore holds that attitude reports with singular 'that'-clauses report attitude states with those same singular propositions as their contents. Consider pairs like (14) and (15) once more.

14. Jameson believes that Spider-Man is a vigilante
15. Jameson believes that Peter Parker is a vigilante

According to the Naïve Russellian, (14) and (15) express the same proposition; they both report that Jameson bears the belief relation to the singular proposition that attributes the property of vigilantism directly to that individual. Consequently, Naïve Russellianism is committed to the counter-intuitive view that substitution of co-referential terms is actually valid. Given that a name is a genuine singular term, its semantic value is exhausted by its referent; thus co-referential names share their semantic values. So, assuming the Compositionality principle that the semantic value of an expression is a function of the semantic values of its component expressions, co-referential names and indexicals must be mutually substitutable in other expressions like sentences.

Thus the Naïve Russellian must explain away the widespread conviction that substitution into small scope positions in attitude reports is *not* valid. As Mark Richard (1990,

p.125) says, it would require "bribery, threats, hypnosis, or the like" to convince most people" that a pair such as (14) and (15) report the same belief.

The characteristic Naïve Russellian rejoinder—from, e.g., from Tom McKay (1981), Nathan Salmon (1986), and Scott Soames (1987, 1988 and 2002)—is that while (14) and (15) semantically express the same proposition, they *pragmatically* convey different propositions. Hearers conflate the pragmatically and semantically conveyed information and take the two reports to express different propositions. (Not all Naïve Russellians offer such pragmatic explanations**.** David Braun (e.g. 1998, 2001a and 2001b) is a notable exception.)

Naïve Russellians have suggested various sorts of information to play this pragmatic role. For example, utterances of (14) and (15) might pragmatically convey the information expressed by (28) and (29). And, since the former is true and the latter is false, hearers may conclude that (14) is true while (15) is false.

28. Jameson would assent to the sentence 'Spider-Man' is a vigilante'
29. Jameson would assent to the sentence 'Peter Parker is a vigilante'

Salmon, Soames, and most other Naïve Russellians have proposed that propositional attitudes are mediated by "ways of apprehending" the proposition, or propositional "guises," and have furthermore proposed that information about the subject's way of taking the proposition can be pragmatically conveyed by an utterance of an attitude report. That is, they distinguish *what* someone believes (a proposition) from *how* he or she believes it (a way of taking the proposition), and they assert that an attitude report *semantically* conveys the "what" and only pragmatically conveys the "how."

For example, Salmon (1986) proposes that while the ordinary language operator 'believes' semantically expresses a dyadic relation between a believer and a proposition, the psychological reality underwriting the truth of the dyadic statements is a triadic relation (he calls it 'BEL') between a believer, a proposition, and a propositional guise. So while (14) and (15) both semantically convey that Jameson believes the same singular proposition, (14) pragmatically conveys that Jameson believes it under 'Spider-Man' ways of thinking, while (15) pragmatically conveys that he believes it under 'Peter-Parker' ways of thinking. And this information—that Jameson believes it under the 'Peter-Parker' ways, really is false; hence the mistaken, though understandable, belief that (15) is false.

The pragmatic solutions offered by Naïve Russellians face significant difficulties. It is implausible that ordinary speakers are normally reasoning by way of meta-linguistic information or information about the believer's ways of taking propositions. There are difficulties as well in spelling out the precise pragmatic mechanism by which the information is allegedly conveyed. For example, Gricean conversational implicature has been proposed—e.g., by Salmon (1986) and by Kirk Ludwig (1996)—but Stephen Schiffer (1987) and others, have raised serious objections for this strategy. The primary problem for Naïve Russellianism remains its counter-intuitive commitment to the validity of substitution.

Hidden Indexical Theories

Hidden Indexical Theories of attitude reports—e.g., from Schiffer (1977), Mark Crimmins and John Perry (1989), and Crimmins (1992a and 1992b)—offer a way of

accommodating the direct reference view that the 'that'-clauses of attitude reports refer to singular propositions while respecting the powerful anti-substitution intuitions. This approach agrees with typical Naïve Russellians that attitudes toward propositions are mediated by ways of apprehending them, but it differs by asserting that information about the subject's way of apprehending a proposition is *semantically*, not just pragmatically, conveyed by an attitude report.

These theories derive their name from the fact that they posit context-sensitive "unarticulated constituents"—i.e., elements of the propositions expressed by attitude reports for which there are no explicit corresponding syntactical elements in the reports themselves—that pertain to the subject's way of taking the proposition. Thus a pair of reports such as (14) and (15) could semantically express different propositions, despite having semantically equivalent 'that'-clauses—e.g., if (14) introduced Jameson's 'Spider-Man' types of ways of apprehending the proposition expressed while (15) introduced his 'Peter-Parker' ways of apprehending it.

Related views have been advanced by Mark Richard (1990), who postulates "Russellian annotated matrices" (the constituents of which are pairs of linguistic entities and their semantic values, such as names paired with their referents) as the contents of attitude states, and by Francois Recanati (1993), who postulates "quasi singular" propositions as contents. The term 'Contextualism' is sometimes used for these sorts of theories, but sometimes instead as a more general name to cover these as well as Hidden Indexical Theories.

Despite accommodating both Direct Reference and anti-substitution intuitions, Hidden Indexical Theories face substantive problems. Bach (1993) and Schiffer (1992) have persuasively argued against the psychological reality of speakers making attitude reports that include ways of apprehending propositions in their truth conditions. Furthermore, the reliance on unarticulated constituents seems *ad hoc*. (See Jason Stanley (2002) for example.) While these theories, like Naïve Russellianism, respect Semantic Innocence, they do so only at the expense of the important principle of Compositionality, since a pair of attitude reports such as (14) and (15) have different semantic value despite all of their proper parts sharing their semantic values.

De Re/De Dicto Distinctions

The cognitive attitudes literature has supposed, often as a load-bearing part of its theories, a distinction between *de re* ("concerning the thing") attitude reports and *de dicto* ("concerning the word/sentence/proposition") attitude reports.

Following Russell, this has usually been drawn as structural distinction in terms of relative scope. In this sense, an attitude report is *de dicto* with respect to the terms (or positions occupied by terms) occurring inside the scope of its cognitive attitude verb, and *de re* with respect to the terms occurring outside that scope. We might call these "structurally *de dicto*" and "structurally *de re*" occurrences.

Let us first consider reports containing descriptions. For example, with respect to 'the dean's partner,' (30) would likely be interpreted as structurally *de dicto* and (31) as structurally *de re*, since the description occurs in small scope in the surface grammar of (30), and in large scope in (31). Furthermore, (31) is structurally *de re* with respect to the anaphoric pronoun 'he.'

30. Otis believes that the dean's partner is a sociologist
31. The dean's partner is such that Otis believes that he is a sociologist

Given Propositionalism and its Relational Analysis, (30) and (31) would have logical forms suggested by the following partial analyses.

32. Believes (Otis, <the dean's partner is a sociologist>)
33. ∃x [x = the dean's partner & Believes (Otis, <X is a sociologist>)]

Thus, structurally *de dicto* occurrences appear to function as part of the characterization of the contents of the reported beliefs, while *de re* occurrences do not. Because the description occurs in small scope in (32), its semantic value is contributed to the proposition expressed by the 'that'-clause and hence to the proposition that Otis is reported to believe. The description in (33), having wide scope, does not contribute its semantic value to the 'that'-clause, and hence not to the reported belief content.

However, although it seems natural to read a sentence like (30) as structurally *de dicto* with respect to the description, matters are more complicated. There is a plausible alternative interpretation of (30) according to which the description does not characterize how Otis is thinking of the individual but only identifies the individual believed to be a sociologist. For instance, if someone overheard Otis saying, "Today's guest speaker is a sociologist," and she knew that the guest speaker was the dean's partner, then she might plausibly report this with an utterance of (30)—even if she thought Otis did *not* know that the speaker was the dean's partner. So it is evident that a "believes-that" sentence such as (30) can be used to express *either* a structurally *de dicto* report or a structurally *de re* report, although in most contexts the former would be more natural, and philosophers of language tend to use "believes-that" sentences as if they were canonically *de dicto*.

Along with an expression's role in content characterization, several other important semantic differences appear to correspond to the structural *de re*/*de dicto* distinction. For example, structurally *de re* occurrences of terms commit the speaker to the existence of a referent, while structurally *de dicto* occurrences do not in general. In the structurally *de re* (31), the speaker is using 'the dean's partner' to refer to the object of Otis's belief, and is thus committed to there being a partner. But, a *de dicto* structure like (32) does not necessarily commit the speaker to the existence of a referent; Otis can bear the belief relation to the proposition that the dean's partner is a sociologist even if the dean has no partner. Similarly, it is not generally valid to apply Existential Generalization (i.e., to "quantify in") to structurally *de dicto* occurrences of descriptions, since from the truth of (32), the truth of (33) does not follow. (Classic discussions of these issues can be found in Quine (1956) and Kaplan (1969), among others.)

Furthermore, since a structurally *de re* occurrence of a term, being outside the 'that'-clause, does not contribute its semantic value to the proposition said to be believed, then any report resulting from replacing the term with a co-referring expression should share the truth value of the original report. That is, substitution of co-referring terms should be valid. Such an occurrence is said to be transparent. A structurally *de dicto* occurrence, being inside the scope of the 'that'-clause, contributes to the belief's propositional content. Therefore, only substitution of a term that shares its semantic value (and not just its referent) can insure preservation of the truth value. That is, substitution can fail. Such an occurrence is said to be opaque.

To summarize: there is an ambiguity involving reports of the form 'X believes that the F is G', which may best be explained as a structural ambiguity between large and small scope occurrences of 'the F'. (Some accounts, e.g. Quine's, instead hold the attitude verbs like 'believes' to be *lexically* ambiguous, positing two distinct belief operators.) With

respect to definite descriptions, at least, the structural distinction is arguably of a piece with several semantic distinctions, at least on the assumptions of Propositionalism.

But only reports containing definite descriptions have been considered so far. Of course, if names or other terms are construed along classic Russellian lines—i.e., as going proxy for descriptions—there is no problem in extending this structural-and-semantic distinction to reports in which names occur. But if ordinary names, indexicals, etc. are construed as genuine singular terms, then the structural *de re/de dicto* distinction and corresponding semantic distinctions come apart from each other. Small scope occurrences of genuine singular terms carry existential commitment, permit EG and permit substitution. In short, they behave rather like large scope descriptions. Consider (34).

34. Daisy believes that Obama is tall

On the Relational Analysis, a structurally *de dicto* reading of (a) attributes to Daisy the property of believing the proposition expressed by 'Obama is tall.' Given direct reference semantics for names, that proposition contains Obama himself a constituent. So the report has existential commitment, since there is no such proposition for Daisy to believe if 'Obama' has no referent. For similar reasons, it appears that Existential Generalization is valid and one may quantify in to the position occupied by the name. For instance, (34) entails (35).

35. $\exists x$ [x = Obama & Believes (Daisy, <X is tall>)]

Finally, since direct reference entails that the co-referential names make the same contribution to propositions expressed (their referent), it follows that two 'that'-clauses like (14) and (15) differing only in substituted co-referential terms must semantically express the same proposition, and—given the Relational Analysis—belief reports with such that clauses must report the same belief. That is, substitution is valid for names. (Of course, this last point is the subject of much debate. The fact that it would license such substitution is frequently the basis of objections direct reference views such as Naïve Russellianism.)

14. Jameson believes that Spider-Man is a vigilante
15. Jameson believes that Peter Parker is a vigilante

Thus there is no univocal answer as to whether a small scope occurrence of a genuine singular term is *de dicto* or *de re*. We might say that such an occurrence is structurally *de dicto*, but semantically *de re*.

Somewhat confusingly, in addition to the *de re/de dicto* distinction among attitude reports, the cognitive attitude literature has also made a metaphysical distinction among attitude *states* using the labels *'de re'* and *'de dicto.'* The distinction is intended to demarcate states that are about specific objects *(de re)* from those that are not *(de dicto)*. For example, in his well-known disambiguation of 'X desires a sloop,' Quine (1956) suggests that on the one ("notional") interpretation it ascribes a desire to own some particular sloop, while on the other ("relational") interpretation it ascribes a desire to own any sloop whatever. The former, object-specific, desire would be a *de re* attitude, while the latter (the general desire to relieve "slooplessness") would be *de dicto*.

The special nature of such *de re* states remains controversial. It has frequently been held to require some particularly close connection between the person and object—e.g., Kaplan's (1969) notion of being causally *en rapport* with an object, or Roderick Chisholm's (1976) notion of "epistemic intimacy." Another common conception of *de re* thoughts is that they are attitude states with a special sort of propositional content, e.g. genuinely singular ("Russellian") propositions. These two approaches can be combined, if one holds that bearing an attitude to a genuinely singular proposition requires a special relation to the singular term's referent.

The two *de re/de dicto* distinctions are sometimes thought to be connected in some systematic way—e.g., that *de dicto* ascriptions can be only be used to ascribe *de dicto* thoughts, while *de re* ascriptions can be employed only to ascribe *de re* thoughts. But there are good reasons for thinking that the two distinctions do not line up so neatly. (See Kent Bach (1987) for a good discussion of such reasons.) Therefore, it is important to be clear about whether with the labels 'de re' and 'dicto' one is characterizing the reports or the reported attitudes. As Bach (1987: 17) noted, "we must be careful to distinguish *de-re*-belief ascriptions from *de re* belief-ascriptions and *de-dicto*-belief ascriptions from *de dicto* belief-ascriptions."

Kripke's Puzzle About Belief

Kripke (1979) introduces an important puzzle that resembles substitution puzzles in some respects, but which does not appear to depend on the assumption of a direct reference account of nouns or involve substitution of terms into attitude reports.

Pierre, initially a French-only speaker, reads a book showing lovely pictures of London captioned as 'Londres' and thus becomes disposed to assent to the sentence 'Londres est jolie.' Kripke concludes, via plausible principles of translation and disquotation (roughly, that assent to sentence S entails belief that S), that Pierre believes that London is pretty. Pierre then moves to a shabby part of London, where he picks up English but somehow has no idea that the city he lives in and knows by the name 'London' is the city called 'Londres.' He is disposed to assent to 'London is *not* pretty' and (again by way of a disquotation principle) Kripke deduces that Pierre believes that London is not pretty. So both of the following appear true, despite ascribing beliefs with contradictory contents.

36. Pierre believes that London is pretty
37. Pierre believes that London is not pretty

But, says Kripke, Pierre cannot have genuinely contradictory beliefs, since anyone is in principle able to detect inconsistent beliefs if he has them, and (as long as he is unaware that the city called 'London' is the one called 'Londres') there is no way that Pierre could possibly recognize that the beliefs in question are inconsistent.

Kripke also constructs a related example that does not involve translation. Peter believes that there is a pianist named 'Paderewski' and that there is a politician named 'Paderewski,' but does not believe they are the same individual. The following reports about Peter are derived from the scenario—another alleged case of contradictory beliefs.

38. Peter believes that Paderewski had musical talent
39. Pierre believes that Paderewski had no musical talent

While Kripke's puzzles must be explained by any account of propositional attitudes, they pose special challenges for those that combine direct reference with the Propositional Theory—for instance, most forms of Naïve Russellianism. Bach (1997) and Shier (1996) have argued that these puzzles are only problematic on the assumption of the Relational Analysis of Attitude Reports. If one does not assume that the contents of beliefs can be "read off" from their 'that'-clauses, then there is no reason to think the pairs above report belief of contradictory contents. (Brian Loar makes a similar point.) Bach (1997: 233) points out that any 'that'-clause can be used, in the right circumstances, to generate a Paderewski case, and uses this as the basis of an argument against the Relational Analysis and its assumption (which he calls "The Specification Assumption) that the 'that'-clause specifies the complete propositional content of a belief.

Anti-individualism and Attitude Reports

Propositional attitudes and attitude reports also figure prominently in the Individualism/Anti-Individualism (Internalism/Externalism) debates in philosophy of mind and language. The influential lines of argument against Individualism about meanings developed by Hilary Putnam (1975), Tyler Burge (1982), and others were extended, by way of Propositionalism, to argue against cognitive individualism.

Cognitive Individualism is the view that a person's beliefs and other propositional attitude states supervene on just her *internal* features and are thus logically independent of the features of the environment. This is often put as the claim that such states are properly individuated by the subject's *narrow* psychological properties where, following Putnam (1975), a narrow property is (roughly) one that "does not presuppose the existence of any individual other than the subject to whom [it] is attributed." According to Cognitive Individualism, any two internally identical persons would share all their attitude states, despite any differences in their environments.

Individualism is intuitively appealing; how could remote external objects and events logically determine what one believes? Certainly they play *causal* roles, but they seem able to do this only in a mediated way by impinging on the *internal* determinants of mental states—i.e., by affecting how the world appears to the individual.

Despite Individualism's intuitive plausibility, it has come under attack from philosophers who hold the semantic thesis of Direct Reference and take it to entail the existence of propositional attitude states with *wide content*—i.e., content that essentially involves contingent external objects. Consider a propositional attitude report containing an occurrence in its 'that'-clause of a directly referential term for a contingent external object.

40. Phyllis believes that a volcano is erupting on Jupiter

Given Direct Reference, this seems to ascribe a belief that essentially involves the planet Jupiter itself—a belief that one simply could not have in a possible world in which Jupiter never existed. (This is also related to one of the common conceptions of *de re* thoughts (above)—that of thoughts with contents that essentially involve particulars.)

The existence of wide beliefs is not entailed by Direct Reference, but by conjoining it with Propositionalism. (See McKinsey (1991 and 1994) for a thorough discussion of this issue.) Given the Relational Analysis of Attitude Reports, (40) reports that Phyllis bears the belief relation the proposition expressed by 'a volcano is erupting on Jupiter.'

And, given the Relational Account of the Attitudes (and the usual further assumption that the propositional content is *individuating* content), Phyllis' belief is type-individuated by that proposition; i.e., one can have the same belief just in case one believes that proposition. Therefore, the belief-individuating proposition will include the planet Jupiter itself, contrary to Individualism.

The recognition that the route from Direct Reference to Cognitive Anti-Individualism runs through Propositionalism suggests a strategy for reconciling Individualism and Direct Reference by challenging Propositionalism, which entails that a report with a wide 'that'-clause must report a wide belief. Some variations of this approach are advanced by Simon Blackburn (1984), McKinsey (e.g. 1991 and 1994), Shier (1996) and Bach (1997).

Related Topics

1.2 Semantics and Pragmatics
1.7 Meaning and Communication
2.1 Reference
2.3 Propositions
3.1 Names
3.6 Anaphora
3.7 Descriptions
3.11 Indexicals and Demonstratives
7.3 Modern Philosophy of Language
7.4 Frege, Russell, and Wittgenstein.

References

Bach, K. (1987) *Thought and Reference*, New York: Oxford University Press.
Bach, K. (1993) "Sometimes a great notion: A critical notice of Mark Crimmins' *Talk about Beliefs*," *Mind and Language* 8: 431–441.
Bach, K. (1997) "Do belief reports report beliefs?," *Pacific Philosophical Quarterly* 78: 215–241.
Blackburn, S. (1984) *Spreading the Word*, Oxford: Oxford University Press.
Braun, D. (1998) "Understanding Belief Reports," *Philosophical Review* 107: 555–595.
Braun, D. (2001a) "Russellianism and Prediction," *Philosophical Studies* 105: 59–105.
Braun, D. (2001b) "Russellianism and Explanation" *Philosophical Perspectives* 15: 253–289.
Burge, T. (1982) "Other Bodies," in A. Woodfield (ed.) *Thought and Object*, Oxford: Oxford University Press.
Carnap, R. (1947) *Meaning and Necessity (2e)*, Chicago: University of Chicago Press.
Chisholm, R. (1976) "Knowledge and Belief: 'De Dicto' and 'De Re'," *Philosophical Studies* 29: 1–20.
Crimmins, M. (1992a) "Context in the Attitudes," *Linguistics and Philosophy* 15: 185–198.
Crimmins, M. (1992b) *Talk about Beliefs*, Cambridge, MA: MIT Press.
Crimmins, M. and J. Perry (1989) "The Prince and the Phone Booth: Reporting Puzzling Beliefs," *Journal of Philosophy* 86: 685–711.
Davidson, D. (1968) "On Saying That," *Synthese* 19: 130–146.
Donnellan, K. (1972) "Proper Names and Identifying Descriptions," *Synthese* 21: 335–358.
Fodor, J. (1981), "Propositional Attitudes" in *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, MA: MIT Press.
Frege, G. (1892) "On Sense and Meaning," in P. Geach and M. Black (eds.) *Translations of the Philosophical Writings of Gottlob Frege* (1952), Oxford: Blackwell.
Kaplan, D. (1969) "Quantifying In," *Synthese* 19: 178–214.
Kaplan, D. (1977) "Demonstratives" in J. Almog, J. Perry, and H. Wettstein (eds.). *Themes from Kaplan* (1989), Oxford: Oxford University Press.

Kripke, S. (1972) *Naming and Necessity*, in D. Davidson and G. Harman (eds.) *Semantics of Natural Language*, Dordrecht: Reidel.
Kripke, S. (1979) "A Puzzle about Belief" in A. Margalit (ed.) *Meaning and Use*, Dordrecht: Reidel.
Loar, B. (1985) "Social Content and Psychological Content," in R. Grimm and D. Merrill (eds.), *Contents of Thoughts*, Tucson: The University of Arizona Press.
Ludwig, K. (1996) "Singular Thought and the Cartesian Theory of Mind," *Noûs* 30: 434–460.
McKay, T. (1981) "On Proper Names in Belief Ascriptions," *Philosophical Studies* 39: 287–303.
McKinsey, M. (1986) "Mental Anaphora," *Synthese* 66: 159–75.
McKinsey, M. (1991) "Anti-Individualism and Privileged Access," *Analysis* 51: 9–16.
McKinsey, M. (1994) "Individuating Beliefs," *Philosophical Perspectives* 8: 303–330.
McKinsey, M. (1999) "The Semantics of Belief Ascriptions," *Noûs* 33: 519–557.
Putnam, H. (1975) "The Meaning of 'Meaning'," in *Philosophical Papers II: Mind, Language, and Reality*, Cambridge: Cambridge University Press.
Quine, W. (1956) "Quantifiers and Propositional Attitudes" *Journal of Philosophy* 53, 177–187.
Recanati, F. (1993) *Direct Reference: From Language to Thought*, Oxford: Blackwell.
Richard, M. (1990) *Propositional Attitudes*, Cambridge: Cambridge University Press.
Russell, B. (1905) "On Denoting" *Mind* 14: 479–493.
Russell, B. (1912) *The Problems of Philosophy*, Oxford: Oxford University Press.
Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.
Saul, J. (2007) *Simple Sentences, Substitution, and Intuitions*, Oxford: Oxford University Press.
Schiffer, S. (1977), "Naming and Knowing," *Midwest Studies in Philosophy* 2: 28–41.
Schiffer, S. (1987) "The 'Fido'-Fido Theory of Belief," *Philosophical Perspectives* 1: 455–480.
Schiffer, S. (1992) "Belief ascription," *Journal of Philosophy* 89: 499–521.
Shier, D. (1996) "Direct Reference for the Narrow-Minded," *Pacific Philosophical Quarterly* 77: 225–248.
Soames, S. (1987) "Direct Reference, Propositional Attitudes, and Semantic Content," *Philosophical Topics* 15: 47–87.
Soames, S. (1988) "Substitutivity," in J. Thomson (ed.) *On Being and Saying*, Cambridge, MA: MIT Press.
Soames, S. (2002) *Beyond Rigidity*, Oxford: Oxford University Press.
Stanley, J. (2000) "Context and logical form" *Linguistics and Philosophy* 23: 391–434.

Section VII

HISTORICAL PERSPECTIVES

7.1
ANCIENT PHILOSOPHY OF LANGUAGE

Luca Castagnoli and Ermelinda Valentina Di Lascio

Ancient philosophers did not introduce a discipline or branch of philosophy called, or straightforwardly identifiable as, "philosophy of language." Nonetheless, they did explore a number of topics which we would classify under this heading. This chapter offers an introduction to the most important of these topics: the nature, origin, and development of language; the relationship between language, thought, and reality; theories of truth; speech-acts; parts of speech and other grammatical issues; linguistic ambiguity. There are a few other relevant topics with which ancient philosophers engaged but which we have not included, not only for space limitations but also because the ancients did not inquire into them in primarily language-related contexts: for example, universals, modalities, logical form, ineffability (for broad introductions cf. Kneale and Kneale 1962 and Mortley 1986). We have adopted a historical perspective, from the Sophists to Augustine, focusing only on the major philosophers and schools, and highlighting thematic and doctrinal continuity and dialectical development.

The Sophists

In the fifth century BC the Sophists emerged as Greece's first professional providers of "higher education"; they were itinerant intellectuals who offered their teaching for large fees to the upper-class youth of the major cities. Although different Sophists specialized in different areas, they all shared a central interest in *logos* ("language"/"speech"/"argument") as a powerful tool within the community. Greeks had always been fascinated by displays of persuasive public speaking and fine oratory to affect political and judicial deliberations or simply to entertain. The rise of Athenian democracy was a fertile ground for increased demand of professional training in public speaking, which the Sophists filled. The most famous Sophist, Protagoras, was notorious for teaching the art of arguing on both sides of any given matter with equal strength, making the weaker argument stronger (cf. also the sophistic-style work *Twofold Arguments*); Gorgias was renowned as a master teacher of rhetoric who could argue persuasively on any matter, regardless of truth and knowledge (cf. *Encomium of Helen*, *Defence of Palamedes*). It is exactly because of their concern with rhetoric that the Sophists engaged in groundbreaking reflection on the nature of language, argument, grammar, etymology

and philology (cf. Guthrie 1971; Kerferd 1981). Unfortunately, our primary evidence is scanty and we rely heavily on limited and fragmentary information from later, and often hostile, sources.

Sophists are reported to have offered teaching on the subjects of "correct diction" (*orthoepeia*) and "correctness of words" (*orthotēs onomatōn*) (cf. e.g. Plato *Cratylus* 384b, 391c). It is not clear whether and how the two subjects differed: the former possibly dealt with the correct use of language in the framework of criticism of poetry. Topics which were certainly included under either heading are: (1) distinction and revisionary discussion of three genders of nouns (masculine, feminine, neuter), credited to Protagoras, who argued that a word's gender should mirror the named thing's natural gender (cf. Aristotle *Sophistical Refutations* 14.173b17–22); (2) distinction of different kinds of speech-acts (command, prayer, narration, address, question, answer), also credited to Protagoras (cf. Aristotle *Poetics* 19.1456b9–19); (3) the phenomenon of synonymy: Prodicus argued for the precise use of words and therefore clarified the different meanings of quasi-synonyms (cf. e.g. Plato *Laches* 197d); (4) the equivocal use of words, discussed by Antiphon (cf. Galen *On Hippocrates' Surgery* 18.2), and exploitation of ambiguity for construction of sophisms; (5) the coinage of neologisms, which should fit exactly the intended concept, again discussed by Antiphon; (6) whether the correct use of words depends on use and convention (*nomos*) or mirrors the nature (*physis*) of the things they are meant to signify (cf. Plato *Cratylus* 384b): (1)–(5) are clearly part of this general discussion and suggest in various ways rejection of the purely conventional hypothesis.

Plato

A key piece of evidence for the Sophists' discussion of (6) is Plato's *Cratylus*, the dialogue most often referred to in analyses of Plato's (ca. 429–347 BC) views on language. Socrates' interlocutors, Cratylus and Hermogenes, offer opposed answers to the question of what makes "names" correct (*onomata* include proper names, nouns, adjectives, participles, infinitives): according to Cratylus a name is correct when it reflects the nature of the nominatum (this position has been labeled "linguistic naturalism"), otherwise it utterly fails to name it; according to Hermogenes, on the contrary, a name is used correctly provided that the use reflects established convention (*sunthēkē*) ("linguistic conventionalism"). Hermogenes' conventionalism is the first to undergo Socrates' refutation. It is paired with a form of Protagorean relativism (things do not have fixed natures, but are relative to each person's judgment) which Hermogenes himself cannot accept (incidentally, linguistic conventionalism is attributed by our sources also to Democritus, whom Plato often paired with Protagoras; cf. Proclus *Commentary on Plato's Cratylus* 16.5–7). Socrates further supports the rejection of conventionalism through examples of etymologies, supposedly showing that our ancestors established and used names as "tools" to mirror a thing's nature: Names are encoded descriptions of things, either by semantic composition of simpler names or, at the basic level of the simplest names, by phonetic imitation of the nature of the nominata ("the primary sounds in a language are vocal imitations of primary properties like fluidity, stability, hardness, largeness, etc." (Sedley 2006: 220)). For example, *eithēr* ("aether") is etymologyzed into *aei thei* ("that which always runs"), which reveals the essence of aether as something in eternal motion (*Cratylus* 410b); in *rhein* ("to flow") the letter *rho* ("r"), produced by vibration of the tongue, imitates motion. Cratylus welcomes Socrates' naturalistic etymologies, but Socrates undermines naturalism as well by arguing that names can never be *perfect*

reproductions of the essences of things, but are often largely incorrect imitations (in these cases the original name-givers showed themselves to be ignorant of the truth). Moreover, some degree of conventional agreement must have occurred in certain cases anyway (for example when a word is composed in equal measure by basic sounds imitating opposite qualities). Therefore, contrary to what Cratylus believes, names are not a good route to the truth. It is debated whether by the end of the *Cratylus* Plato himself remains committed to a definite theory about "correctness of names." Most scholars (for example Williams 1981) have interpreted the conclusion as a complete rejection of naturalism, and thus possibly as an endorsement of a moderate form of conventionalism (as Crivelli 2008 suggests, in modern terms Plato would reject at least the view that all names function as definite descriptions used attributively). Others (especially Sedley 2003) have suggested, on the contrary, that Plato did accept the naturalism defended in the central part of the dialogue, even if mitigated by elements of conventionalism introduced later. However that might be, in the close of the *Cratylus* Plato himself issues the warning that the study of language cannot be a privileged means to pursue knowledge of reality; we should instead inquire directly into the things themselves, i.e. the Forms.

This warning might appear surprising in light of two claims often made by scholars: (1) names are, primarily, names of the Forms and only indirectly of the things which participate in the Forms and are therefore named after them (cf. *Phaedo* 102b1–2; *Republic* 10.596a6–7); (2) to understand the meaning of the word "F" is to grasp the Form of F. However, even if we grant that "there is a linguistic dimension to Plato's theory of forms" (Crivelli 2008: 219), Plato's inquiry into the Forms is not driven by an interest in the workings of language, nor are its results used to draw definite, explicit conclusions concerning the semantic relation between language and reality (we might still ask how, in our terms, Plato would have spelled out that relation). For example, the definitional inquiry typical of Plato's dialectical method ("What is F?") is not after the referent or sense of a *word* ("F"), whether "correct" or as determined by use, but after the essence of a *thing* (the F), although Plato might of course have recognized that determining the latter had important dividends for identifying or clarifying the former. Incidentally, Plato's definitional inquiries in most dialogues are narrowly focused, especially on moral qualities (for example courage in the *Laches*, temperance in the *Charmides*, justice in the *Republic*). This limitation is also reflected in the illustrative lists of Forms which occur in many passages and in the denial that there are Forms corresponding to all general terms in other passages (for example *Parmenides* 130b–e; *Statesman* 262a–263b). Therefore, if we took Forms as part of a theory of meaning, we would end up admitting that for Plato not all general terms had meaning.

Plato's limited attention to the phenomena of synonymy and homonymy confirms his secondary interest in language *per se*. In the dialogues, Socrates often makes fun of Prodicus' obsession with semantic distinctions of quasi-synonyms (for example *Protagoras* 340b–341d, 358a–b), and when dealing with sophisms in the *Euthydemus* he engages in homonymy disambiguation only once (277e–278a), even if he is clearly aware that many sophisms exploit linguistic anomalies.

We are not suggesting, of course, that Plato was not interested in language at all. The philosophical method to achieve knowledge of the Forms, "dialectic," consists in reasoned assessment of one's own and others' beliefs through question-and-answer exchanges. That such exchanges can only be carried out through a (shared) language is no accident; more than once Plato expresses the idea that human thinking and judging are themselves a sort of inner silent dialogue with oneself (*Theaetetus* 189e–190a;

Sophist 263e–264b; *Philebus* 38c–e; whether for Plato there are *also* higher forms of non-discursive cognition is a much debated question beyond the scope of this chapter). Exactly because of this role of language in philosophical inquiry Plato was aware of the risk that linguistic errors could engender philosophical errors. The main example is Plato's treatment of the puzzle of how it is possible to speak (and judge) falsely. The puzzle, which originated with certain readings of Parmenides (fr. 2.7–8: "neither could you know what is not—for it is impossible—nor could you mention it") and is attributed by Plato to Protagoras, is presented in the dialogues in various forms (*Cratylus* 428e–430d; *Euthydemus* 283b–288a; *Republic* 478b–c; *Theaetetus* 188c–189b; *Sophist* 261b–264b; cf. Denyer 1991). In its simplest formulation the puzzle can be schematized as follows: to speak falsely is to say what is not; to say what is not is not to say anything; therefore one cannot speak falsely (one either speaks the truth or does not speak at all; in the *Cratylus* the latter is compared to making meaningless noises). Plato's solution, presented in the *Sophist*, consists in clarifying that making a "statement" (*logos*) is something complex and thus different from naming something or "saying what absolutely is not." Even the simplest statement—the bearer of truth and falsehood—is a combination of a "noun/subject" (*onoma*) and a "verb/predicate" (*rhēma*), in which the latter states something of the named object. Therefore to make a false statement ("Theaetetus flies") is not to name a nonexistent object (flying Theaetetus), but to say of something that exists (Theaetetus) something which also exists (flying) but does not "combine" with it. Conversely, a true statement expresses a *real* combination between the named object and what is said of it (in the *Theaetetus* Plato rejected a form of extreme Protagorean relativism according to which truth depends on individual judgment). Plato's pioneering distinction between the two essential components of a statement is a fundamental contribution. It is not clear, however, whether it amounts to a grammatical distinction (noun/verb), a syntactical one (subject/predicate), or an ontological one (thing/property).

Aristotle

Plato's distinction in the *Sophist* was expanded upon by his pupil Aristotle (ca. 384–322 BC). Although his views on the matter are expressed only in embryonic and scattered form, Aristotle was the first to define important phonetic, grammatical and syntactical elements. In *Poetics* 20 he distinguishes:

- three nonsignificant linguistic items: letter, syllable, "connector" (*sundesmos* includes particles, conjunctions, prepositions);
- four significant (*sēmantikai*) linguistic items: "name" (*onoma* includes proper names, nouns, adjectives, pronouns, participles), "verb" (*rhēma*), "modification" of a name or a verb (*ptōsis* includes what we call "case" and "inflection"), "phrase"/"sentence" (*logos*).

Language is distinctive of human beings (and indeed what makes us human: cf. *Politics* 1.2). While animal cries are not articulate, human speech is the articulation of the voice into letters, the basic elements out of which compound sounds (syllables, words, phrases, and sentences) are composed.

Obviously Aristotle's views on *significant* linguistic items are especially relevant here. These are discussed also in *On Interpretation*, whose opening offers a general view of significant language:

> spoken sounds are symbols (*sumbola*) of affections (*pathēmata*) in the soul, and written marks symbols of spoken sounds. And just as written marks are not the same for all men, neither are spoken sounds. But what these are in the first place signs of–affections of the soul–are the same for all; and what these affections are likenesses of–things–are also the same.
>
> (1.16a1–7)

Language is conventional, as suggested by the fact that the same thoughts ("affections in the soul") are expressed through different vocalizations and, indirectly, different written marks in different languages. Oral and written words are adopted primarily as "symbols" or "tokens" and "signs" of thoughts and, indirectly, of things (cf. *Sophistical Refutations* 1 for an alternative formulation which makes the symbolic link between words and things direct). For example, "a name is significant (*sēmantikē*) 'by convention (*kata sunthēkēn*)' because no name is a name by nature, but only when it becomes a symbol" (*On Interpretation* 2.16a29–30). The term *sumbolon* is particularly well suited to the idea of conventionality, since:

> the normal use of the Greek word was for a tally or token. A contract or other agreement might be marked by breaking a knucklebone or other objects in two, one portion being taken by each of the parties to the agreement. Each person kept his piece, and could identify the person who presented the other piece by matching it with his own . . . the meaning of a word is fixed by convention . . . just as the importance attached to a tally, token, or ticket depends on agreement between the parties concerned.
>
> (Whitaker 1996: 10)

The relation between thought and the world is, on the contrary, natural. Thoughts are the same for all human beings because so are the things they represent: literally, a thought is a "likeness" of a thing, because it shares the "form" of the thing thought (for this aspect of Aristotle's philosophy of mind cf. *On the Soul*). Aristotle seems to engage with the aporetic end of Plato's *Cratylus* by endorsing linguistic conventionalism and transferring the naturalist element from the language-world relation to the thought-world relation (Aristotle's terminology alludes to the *Cratylus*: cf. for example "every *logos* is significant not as a *tool*, but by convention" at *On Interpretation* 4.17a1–2).

Conventionalism is also the starting point of Aristotle's specific discussion of the distinctive features of names, verbs (and their modifications) and phrases/sentences. All these are first defined as "conventionally significant sounds." Names and verbs, albeit composed of letters and syllables, have no part significant by themselves, whereas phrases/sentences do (the component names, verbs, and modifications). Verbs in turn differ from names in that they "signify time in addition" and are "signs of things predicated of something else" (*On Interpretation* 3.16b6–8); when they are taken in isolation, and thus lack predicative force, verbs are names. Names need not fill the syntactic role of subjects ("man is an animal"), but they too can be used as predicates when connected to the subject through a copula expressing the temporal and predicative verbal functions ("Socrates is a man").

Names, verbs, and those *logoi* (phrases) which are composed only of names (for example a *definiens* of "man" such as "two-footed terrestrial animal") are neither true nor false. Sentences (*logoi*) are the result of the combination of names and verbs or of names and names through a copula. Among sentences, the "assertoric" (*apophantikoi*) ones are

the bearers of truth and falsehood (and so are the corresponding thoughts), whereas other kinds of sentences (for example prayers) cannot be true or false. The most basic assertoric sentences are affirmations and negations. In affirmations, something is said to be combined with something else; in negations something is said to be divided from something else. Affirmations are then true when the things which are said to be combined are in fact combined in the world, and false when they are actually divided; negations are true when the things which are said to be divided are in fact divided, and false when they are actually combined. In other words, Aristotle adopted a correspondence theory of truth: Truth consists in the correct match of combination/division in thought and speech with combination/division in reality (cf. also *Metaphysics* 4.7.1011b25–7: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is and what is not that it is not, is true"). Affirmations having empty names as subjects are always false, while the corresponding negations are always true (cf. *Categories* 10.13b27–35).

One might wonder whether Aristotle's views on the "signification" relationship between language, thought, and reality can be construed as a definite theory of meaning. This is a controversial issue, and any reconstruction runs the risk of over-interpreting Aristotle's scattered remarks in light of the conceptual apparatus of modern philosophy of language. The general picture emerging from some modern readings of Aristotle's texts can be sketched as follows. Words signify (*sēmainein*): (1) our common thoughts associated with them ("thunder" signifies the thought of a noise accompanied by lightning) and (2) essences of things in the world ("thunder" signifies the essence of thunder, i.e. noise in the clouds caused by quenching of fire). Words to which no essence corresponds, such as empty terms (for example "goat-stag") and indefinite terms (for example "non-man"), still have signification in sense (1). All remaining words signify in both senses, even when the speaker and the hearer lack (full) knowledge of the essence signified. One main point wherein modern accounts of Aristotle's theory differ is in the understanding of which sense of "signification" is primary, and thus of whether Aristotle's theory of signification is a theory of meaning at all. For example, Irwin 1981 stresses the primacy of essential signification (what he calls "signification by nature") and therefore denies that Aristotle had a theory of meaning, where by "meaning" of a word Irwin intends "what is entered in a dictionary; . . . what a competent speaker and hearer grasp but a non-speaker does not grasp; what two synonymous words share" (242). (This is mere "signification to us," in which Aristotle was not interested except as the starting point of the inquiry into essences, i.e. primary signification.) Other scholars (for example Charles 1994; Modrak 2001), on the contrary, place more weight on the direct "symbolic" and "semantic" relationship between words and thoughts described in *On Intepretation*, i.e. signification in sense (1). This is the primary signification of words and therefore Aristotle does have a theory of meaning, provided that meaning is not identified with Fregean sense or reference but with intentional content of thoughts.

What makes attempts to reconstruct Aristotle's *theory* of language all the more difficult is the fact that, just as in Plato's case, Aristotle's relevant remarks belong to different contexts in which the primary interest is not language *per se*. The *Poetics* account of significant and nonsignificant linguistic items is linked to the broader discussion of poetical composition and literary criticism. In *On Interpretation* Aristotle's discussion is guided by his interest in dialectical arguments. For example "his analysis of the structure of the assertion is meant to apply to the ideal assertions prescribed in dialectic and not to everyday, idiomatic utterances" (Whitaker 1996: 59). In *Metaphysics* 4, the concept

of signification is central to Aristotle's renowned defense of the Principle of Noncontradiction: if communication is to be possible, words must bear definite signification, but then the Principle of Noncontradiction must hold good ("X" cannot both signify and not signify X).

Another instrumental discussion of language occurs in the *Sophistical Refutations*, where Aristotle classifies sophisms on the basis of the fallacy involved, with a view to offering a method for solving them so as to avoid defeat in dialectical exchanges (this classification is still influential in modern treatments of fallacies). Six of the thirteen kinds of sophisms identified are based on *linguistic* fallacies which deceptively exploit certain anomalies of language: (1) *homonymous* words signify more than one thing; (2) *amphibolous* sentences can be syntactically construed in two semantically different ways; (3) parts of a sentence can be *combined* or (4) *divided* so that the whole has, accordingly, different significations; (5) words differing only in their "*intonation*" (i.e. identically written but pronounced differently) do not signify the same; (6) words having the same "*form of expression*," i.e. belonging to the same grammatical category or playing the same syntactical role in a sentence, can signify different kinds of things ("Socrates" and "man" are both names which can function as subjects, but the former signifies an individual, the latter a universal).

Other discussions of the phenomenon of homonymy occur elsewhere within Aristotle's criticism of rival philosophical views (for example the rejection of the Platonic univocity of goodness; cf. *Nichomachean Ethics* 1.6) and defense of his own theories (one notorious case is that of the homonymy of being; on the various forms of homonymy cf. Shields 1999).

Hellenistic Philosophy

Inquiries into various aspects of language (for example its origin, grammar, syntax, semantics) became more focused and systematic during the Hellenistic age (the three centuries after Aristotle's death). However, our extant evidence is unfortunately scanty: most of the vast output of the four major schools (Epicureans, Stoics, Academics, Peripatetics) is lost, and often we have to rely on a few short surviving fragments or on much later reports and paraphrases, second or thirdhand and only partially reliable.

Epicureans

Epicurus' (341–270 BC) position on the origin of language can be interpreted as a reaction to Aristotle's main argument for linguistic conventionalism (Sedley 1973: 20). Aristotle had inferred the conventionality of language from the fact that mental affections (and the things in the world producing them) are the same for all human beings but words differ in different languages; Epicurus, on the contrary, takes the existence of different languages as evidence for the natural, nonconventional origin of language. In a dense section of his *Letter to Herodotus* (75–76) Epicurus first outlines a general view of the origin and development of civilization, articulated into a first phase dominated by nature and a later one in which human reason perfected nature's contribution. Then he applies this view to the paradigmatic case of language, also in two phases. In the first natural phase men were instinctively led to make utterances by the "impressions" (*phantasmata*) and "affections" (*pathē*) generated in them through interaction with the external world (just as animals do). Probably the original utterances were already articulate

to some degree (Epicurus calls them "names"). Since different objects caused different impressions and affections, which in turn were naturally expressed through different utterances, different words came to be used for different objects. However, environmental factors and, possibly, physiological ethnic variations contributed to the formation of different psychological states in human beings living in different places. This natural difference explains why different human groups naturally came to use different words for the same thing, i.e. the variety of natural languages (which thus supports linguistic naturalism). This sketch of the first phase of the origin of language would be accepted by most scholars (for example Verlinsky 2005). According to an alternative reconstruction originally utterances differed for each individual, because so did impressions and affections, and thus each individual developed a *private language*, whereas only in the second phase a common language was adopted by conventional imposition. Brunschwig 1994 has convincingly rejected this private language hypothesis; this, however, does not mean that for the Epicureans thoughts cannot exist independently of public language, and the Epicurean theory seems in fact to be committed to the logical and psychological priority of thought, at least in its basic forms, over language. The naturalistic view of the origin of language is opposed by Epicurus to a conventionalist one: "names did not originally come into being by imposition (*thesei*)." The allusion here (and in the parallel passage in Lucretius *On Nature* 5.1028–1090) seems to be to a human inventor (*contra* the *Cratylus*' name-giver hypothesis), but in a later Epicurean source (Diogenes of Oenoanda 10.2) the possibility of a divine inventor of language is also discarded and ridiculed.

In the second phase of language development, reasoning and *convention* intervened: Natural languages were refined ("particular words were imposed," Epicurus says, with the purpose of making language less equivocal and more concise) and enriched ("the vocabulary was artificially enlarged, or at any rate existing words were given additional meanings, by intellectuals turning their attention to abstruse or theoretical entities" (Long and Sedley 1987: 100)). Thus a novelty of Epicurus' proposal consists in combining a naturalistic theory of the origin of language with an explanation of the fundamental (albeit nonprimary) role played by convention (notice, however, that Epicurus' naturalism is very different from Cratylean naturalism, according to which words should mirror the nature of things).

Whereas with their views on the origin and nature of language the Epicureans made a clear contribution to a well-established tradition, the question of whether they had a definite theory of meaning is more difficult to answer, just as it was in Aristotle's case. Some interpreters (e.g. Atherton 2009) suggest that it would be anachronistic to project onto the few extant relevant texts any semantic theory, and that the Epicureans were interested in the phenomenon of linguistic communication rather than our question of meaning. Others believe that the Epicureans were committed to an "extensionalist" semantics according to which a direct signifying relation holds between linguistic items and things in the world. Most scholars suggest that, on the contrary, the Epicureans were "intensionalists" and some fundamental semantic role was played by intermediate items between language and things. According to the latter interpreters this intermediate role was played by mental items, in a way somewhat analogous to that of Aristotle's "affections." In his *Letter to Herodotus* (37) Epicurus writes that our "primary concepts" "correspond" to our words and "underlie" them, and that without them our utterances would be "empty" and no inquiry would be possible; Diogenes Laertius (10.33) reports that for the Epicureans "we would not have named something if we had not previously

learned its delineation by means of a preconception (*prolēpsis*)." "Primary concepts" or "preconceptions" are in Epicurean jargon the fundamental, naturally formed and commonly shared universal concepts which arise in the human mind through repeated experience and memory (for example the concept of horse derived since childhood from repeated encounters with horses). Intentionalist interpreters believe that these basic concepts are themselves meanings (Long and Sedley 1987: 101) or, at least, that words have their meanings in virtue of standing in some relevant relation to these psychological states of the speakers (cf. Everson 1994). These suggestions are compatible with the texts mentioned above. However, given the epistemological focus of the context of those texts, an intentionalist interpretation is by no means required, just as it is not required by Epicurus' account of glossogenesis (mental states are described as what prompted natural vocalizations, but not as the meanings of those vocalizations). The strongest evidence for the rival extensionalist interpretation lies in a couple of later reports (Plutarch *Against Colotes* 1119F–1120A; Sextus Empiricus *Against the Logicians* 2.11–12) according to which, of the Stoic triad words—"sayables"—things (cf. p. 820), the Epicureans rejected the intermediate class of "sayables," thus leaving words as truth-bearers and things themselves as their meanings. It has been noted, however, that the Epicurean rejection of "sayables" does not imply that no alternative intermediary semantic role could be played by some other item, such as preconceptions. Our sources possibly drew inferences about the Epicurean position rather than report it. The situation is further complicated by the possible distinction between word-meaning (which might have been extensional) and speaker-meaning, in which the speaker's mental state could have played a role (cf. Barnes 1999: 195–7); however, the hypothesis that the Epicureans themselves might have been perceptive to such a distinction is, again, conjectural. It seems safe to conclude that either the Epicureans had no interest in semantic theory at all or the evidence is anyway too scanty for us to adjudicate between intensionalist and extensionalist interpretations.

Stoics

Whereas the Epicureans had little interest in grammar, syntax, and linguistic phenomena such as ambiguity, all these topics were object of intense attention within the Stoic school, which was founded by Zeno of Citium toward the end of the fourth century BC and flourished until the end of antiquity. Judging from our limited evidence, it is not clear whether the Stoics also held a full-blown theory on the origin and nature of language. According to Origen (*Against Celsus* 1.24), the Stoics thought that names are natural and the first human vocalizations imitated things. Augustine (*On Dialectic* 6) explains that they believed that language originated with onomatopoeic imitation and different principles intervened subsequently in the formation of other words (presumably the Stoics used these principles in the etymological work in which they engaged extensively). The idea that first sounds mirrored things, however, differed from Cratylean naturalism: For the Stoics only the first words were mimetic, and they were so onomatopoeically (because of the high communicative power of onomatopoeia; cf. Allen 2005), and not by mirroring the essences of their nominata (even so, etymologies were considered worth studying as means of information about our ancestors' understanding of reality).

Chrisippus (280–207 BC) was the most original exponent of Stoicism, especially in the broad area of philosophy labeled "logic"; he wrote at least 130 works on a variety

of logical subjects, including many which could be catalogued under the heading "philosophy of language" (virtually all these works have been lost). These subjects were studied by the Stoics as part of "dialectic," one of the main branches of "logic," which encompassed everything that had to do with *logos* ("rational discourse"). According to Chrysippus, the subject matter of dialectic was "(1) signifiers and (2) things signified" (Diogenes Laertius 7.62): (1) utterances, expressions, parts of speech, dialects, virtues and vices of speech, ambiguity, euphony, poetry; (2) incomplete and complete "sayables" (*lekta*), truth, modalities, arguments, validity and invalidity, paradoxes, and sophisms. This main subdivision already suggests that the concept of signification was central to the Stoic logical and linguistic theory.

The realm of signifiers includes physical ("corporeal") entities: at the most generic level, "utterance" or "sound" (*phōnē*) is simply air that has been struck and becomes the object of hearing. Human beings are capable of articulate and alphabetic utterances, i.e. "expressions" (*lexeis*); but only signifying expressions, issuing from thought, count as "speech" (*logos*). The Stoics classified various "parts of speech": "proper name" (*onoma*), "appellative" (*prosēgoria* includes nouns and adjectives), "verb" (*rhēma*), "connector" (*syndesmos*), "article," and "adverb." Although the Stoics were interested in phonetic and purely grammatical aspects of language (for example dialects, defined as sets of *lexeis* encompassing the distinctive phonetic practices of a linguistic community, and solecism, studied among the vices of speech) and were hugely influential upon the development of Greco-Roman grammar, it is important to note that the parts of speech were uniquely defined on the basis of what they signified or their syntactic contribution to signification. For example, a proper name is that part of speech which signifies a corporeal "peculiar quality" (for example "Socrates" signifies the quality of "Socratesness:" according to the Stoics, each individual has a quality which makes it distinct from everything else). An appellative is that part of speech which signifies a corporeal "common quality" (for example "horse" signifies the quality of "horseness"). A verb is that part of speech which signifies an incorporeal "predicable" (for example "write" signifies writing) and a connector is that part of speech which binds other parts of speech together.

Central to the Stoic theory of language, and foundational for the part on "things signified," is the concept of "sayable" (*lekton*). When we speak, we *utter* significant expressions in a certain language, but what thereby gets *said* and signified are incorporeal sayables, the meanings of our words that are not understood by those who do not speak that language (a sayable is, literally, "something which can be said" by uttering a certain expression). Since in Stoic ontology only what is corporeal exists, incorporeal sayables (also called *pragmata*, "things" or "states of affairs") cannot exist but "subsist" (like the other incorporeals —space, void, and time). The Stoics thus introduced a three-tiered account of signification, according to which: (1) linguistic expressions (corporeal signifiers) signify (2) incorporeal *lekta*, what gets said through those expressions *about* (3) corporeal objects in the world and their corporeal qualities. For example, (1) "Dion walks" signifies (2) that Dion walks about (3) Dion (cf. Sextus Empiricus *Against the Logicians* 2.11–12; Seneca *Letters* 117.13). This account has often been associated with the Fregean distinctions between word, sense, and reference (cf. for example Mates 1961: 19–26), but the parallelism is much less straightforward than it has been maintained (cf. Frede 1994). The role of sayables as significations of words, distinct from the things in the world of which we speak, can also be compared to the intermediary role which mental "affections" play in Aristotle's account of signification (and which "preconceptions" might have played in the Epicurean theory). Whether or not Stoic

lekta were introduced to play the same structural role as Aristotelian *pathēmata*, they are not mental objects but public entities which secure the possibility of communication: It is the same sayable that is expressed by a certain speaker and becomes the content of a listener's thought.

Although sayables are not thoughts (which for the Stoics are corporeal), they function as the intentional content of our thoughts (cf. Sextus Empiricus *Against the Logicians* 2.70: "a sayable is what subsists in accordance with a rational impression"). Human beings differ from other animals exactly in that their "impressions" are "rational" (*logikai phantasiai*), i.e. can be shown forth through significant speech (*logos*). Whereas some animals, like parrots, have also "uttered speech" (i.e. produce articulate sounds), only human beings have "inner speech" (*endiathetos logos*) which is a function of rationality (Sextus Empiricus *Against the Logicians* 2.275–6). Thus a strict relationship between language and thought holds: There is nothing sayable that cannot be thought, and nothing which we can think that cannot be said (cf. Barnes 1999: 213). And since *logos* was also, for the Stoics, the divine organizing rational principle of the universe, the relationship extends to reality. Reality is accessible by thought and language, and what can be said by rational beings is isomorphic with the rational structure of reality.

The origin of the Stoic *lekton* and the rationale for its incorporeal nature and subsistence are much debated (cf. Long and Sedley 1987: 199–202; Frede 1994; Barnes 1999). Similarly debated is whether sayables are completely mind-independent entities or whether, while being clearly distinct from signifiers and thoughts, they exist only as long as the words or thoughts expressing them do (cf. Frede 1994: 118 and Long 1971: 97 for the two positions). The technical details of the theory are also unclear. For example, although the majority of sources suggest that proper names and appellatives do not signify *lekta* but corporeal qualities of objects (cf. Diogenes Laertius 7.58), a case can be made for them signifying *lekta*, despite the fact that "predicables" (*katēgorēmata*), signified by verbs, are the only examples routinely used for "incomplete sayables" in our sources. (The Stoics wrote extensively about predicables and classified them: cf. Barnes 1999: 204–6.) The signification of a verb is fixed by what, by uttering it, you can say *about* something: For example, "to run" signifies something (running) which can be said *of something* (e.g. Socrates). This predicable, running, is an incomplete sayable because when it gets expressed (for example by the English "runs") the hearers feel prompted to ask for a completion ("who runs?"), and no complete state of affairs gets expressed until the predicable is attached to a "nominative case" (cf. Diogenes Laertius 7.63–4). The notion of "case" is itself controversial (are cases signified by proper names/appellatives, and if so what is their relationship with peculiar/common qualities? Are cases incomplete *lekta*?), interestingly straying between grammar, semantics and metaphysics.

"Complete" sayables were classified on the basis of what you *do* by saying them: for example "assertibles" (*axiōmata*) are sayables "saying which we make assertions, and which are either true or false," while imperatival sayables are "those saying which we give a command" (and which are neither true nor false) (Sextus Empiricus *Against the Logicians* 2.71–3). The Stoics distinguished other kinds of complete sayables, including questions (which require "yes" or "no" as an answer), inquiries (open questions requiring a discursive answer), puzzlements (questions which do not require answer), oaths, prayers, invocations, curses, hypotheses and expositions (for example "let this be a straight line"). The Stoic classification of sayables thus resembles modern classifications of speech-acts: Each time we express a complete sayable we (1) utter a linguistic expression, (2) say the associated sayable and, (3) perform a certain speech-act. Unlike

Aristotle, who believed that nonassertoric speech belongs to the realm of rhetoric and poetics, for the Stoics the study of all types of sayables belonged to dialectic.

Stoic assertibles have often been associated with propositions: They are the primary bearers of truth and falsehood (assertoric sentences and impressions are true or false derivatively, in virtue of having assertibles as their propositional contents). The Stoics adopted a correspondence theory of truth: "A true assertible is that which is the case and is contradictory of something, and a falsehood is that which is not the case and is contradictory of something" (Sextus Empiricus *Against the Logicians* 2.85). Although no assertible exists, then, within subsisting assertibles, the Stoics drew a distinction between true assertibles, which "are the case" or "obtain," and false assertibles, which are not the case. They also subscribed to a "temporalized" version of the principle of bivalence: Assertibles are either true or false *at times*, and can change their truth-values over time. For example, the assertible that it is raining is true now, because it is raining, but will be false in one hour, when it will not be raining. Stoic assertibles also differ from propositions in that they are tensed and can contain token-reflexive elements, such as "I" or "this." The sentence "I am now eating this," uttered at different times by different speakers pointing at different objects, signifies the very same assertible but different propositions. Incidentally, assertibles are also bearers of modalities (possibility, impossibility, necessity) of which Chrysippus advanced the discussion, initiated by Aristotle and continued by the Megarians Diodorus Cronus and Philo in the second half of the fourth and first half of the third century BC.

The Stoics embarked upon a sophisticated classification of assertibles, both "simple" and "non-simple." Among simple assertibles they distinguished (cf. Sextus Empiricus *Against the Logicians* 2.96–100; Diogenes Laertius 7.69–70):

(1) "definite" assertibles, involving some deictic reference (for example "*this one* is walking"; the pronouns "I" and "you" were also considered deictic). Definite assertibles are "destroyed" when the object of the deictic reference no longer exists (for example "this one is dead," where "this one" refers to Dion, is false as long as Dion is alive, and is destroyed, i.e. ceases to subsist, at Dion's death);
(2) "intermediate" assertibles, composed of a nominative case and a predicable (for example "*Dion* is dead," which is false as long as Dion is alive, and becomes true from the moment of Dion's death);
(3) "indefinite" assertibles, governed by an "indefinite particle" (for example "*someone* is walking"). An indefinite assertible is true whenever a corresponding definite one (for example "this one is walking") is true.

Negations of simple assertibles count as simple assertibles. The Stoics distinguished three forms of negative assertibles:

(1) "negations," formed by prefixing the particle "not" (for example "*Not*: it is day"). External negation is truth-functional: If an assertible *p* is true, then its negation not-*p* is false; if *p* is false, then not-*p* is true. An assertible and its negation are contradictory, and the Principle of Excluded Middle holds good;
(2) "denials," having a compound negative as subject (for example "*no-one* is walking");
(3) "privations," having a privative particle prefixed to the predicable (for example "this one is *un*-kind").

Interestingly, ordinary negative sentences, in which the negation governs the predicate (for example "Dion is *not* walking"), were not classified under this scheme, and indeed counted as affirmations for the Stoics, who insisted that in "negations" the negative particle should have broad scope, governing the whole assertible. Not only does the external negation allow recognition of the quality of the sentence at first sight, but it also guarantees the principle that if an assertible is true (false) then its negation is false (true) and guarantees the Principle of Excluded Middle. When the subject X is empty, both "X is F" and "X is not F" are false, because both assertibles have existential import (for example if Dion does not exist, both "Dion is walking" and "Dion is not walking" are false), whereas "Not: X is F" ("Not: Dion is walking") is true. This shows that the classification of assertibles was not guided merely by observation of existing linguistic structures but by logical or metaphysical considerations.

This is even clearer in the classification of nonsimple assertibles, which are composed of two or more simple assertibles (or the same assertible taken two or more times) linked by some connector. The Stoics made the largest contribution to propositional logic in antiquity (although the Peripatetic Theophrastus and the Megarians had initiated study of the subject), and nonsimple assertibles are the building blocks of that logic. The study of nonsimple assertibles goes beyond the scope of this chapter, and belongs more properly to the field of philosophy of logic (for an overview cf. Mates 1961; Bobzien 1999). We only mention here that the Stoics distinguished conjunctions, negated conjunctions, disjunctions, para-disjunctions, conditionals, para-conditionals, causals, and comparatives, regimenting the natural language usage of these assertibles (for example connectors are always prefixed) and providing analyses of their truth-conditions (never truth-functional, except for conjunctions).

As mentioned above, within their dialectic the Stoics investigated linguistic ambiguity (cf. Atherton 1993). Chrysippus held that every expression is ambiguous. Supposedly he meant that each word signifies both its significate and itself, hinting at something analogous to our use-mention distinction. He also insisted on the lexical character of ambiguity, in opposition to the claim of Diodorus Cronus that ambiguity does not exist because the meaning of a word or a sentence is the meaning intended by the speaker (this is connected to Diodorus' own extreme conventionalism). We also inherited a (possibly Chrysippean) classification of eight types of ambiguities (cf. Galen *On Linguistic Sophisms* 4).

Although the Stoics do not seem to have discussed the Sorites paradox in terms of lexical ambiguity, they devoted great attention to the Sorites, which had been devised by the Megarian Eubulides and used by the Stoics' skeptical adversaries to undermine their epistemological optimism. The reconstruction of Chrysippus' diagnosis and solution is deeply controversial. Several interpreters have attributed to him an epistemicist position according to which, although it is impossible for us to determine which of the conditionals involved is false (for example what precise cut-off value of x falsifies the conditional "if x grains of sand are not a heap, x+1 grains of sand are not a heap either"), there is, unbeknown to us, one false conditional in the argument, which is thereby materially false (cf. Barnes 1984). Other interpreters have rejected this reconstruction, suggesting that for Chrysippus some of the borderline cases have no truth-value (possibly because they fail to express assertibles on account of vagueness; cf. Bobzien 2002).

Although no part of Stoic philosophy was labeled "grammar" or was clearly identifiable with grammar, there is evidence that the Stoics initiated the study of our

"traditional grammar" (which includes phonology, morphology, syntax, parts of speech and their features), labeled "technical grammar" in the Hellenistic period. (It is often assumed that the Alexandrian school initiated this study, but actually the school's main interests were philology and literary criticism.) Not only is the classification of parts of speech obviously relevant to grammar, but also the discussion of the way in which simple and compound *lekta* are structured contributed to the development of syntax in later antiquity (Dionysius Thrax was the first to write about technical grammar between the first and second century BC, while Apollonius Dyscolus authored the first extant treatise on syntax in the second century AD; cf. Frede 1987).

Later Antiquity

The development of grammar is by no means the only example of sustained interest in language in late antiquity. Peripatetics and Platonists reflected on the views on language proposed in the Classical and Hellenistic periods, while Pyrrhonian skeptics attacked the whole previous philosophical tradition, including theories of language (cf. Sextus Empiricus *Outlines of Pyrrhonism* 2; *Against the Logicians*; *Against the Grammarians*), and devised at the same time nonassertoric modes of speech to express their skeptical stance. We can only mention here the closing chapter in this later history, which falls outside the Peripatetic, Platonist, and Pyrrhonian traditions.

Augustine's (354–430 AD) insights on language are mostly known to students of philosophy through the medium of the influential presentation and criticism of the "Augustinian picture of language" which Wittgenstein sketched in the *Brown Book* and the *Philosophical Investigations*. According to that picture, individual words name objects, and the meaning of each word is the object for which it stands. Some scholars (e.g. Burnyeat 1987) have observed that this picture does not really reflect Augustine's historical view. It is true that Augustine wrote that infants begin learning language ostensively by interpreting adults' behavior and picking up the names which they observe get associated with things (*Confessions* 1.6). There is no indication, however, that *all* language acquisition works in this way (Augustine was aware, anyway, that ostension is chronically and unavoidably ambiguous as a way of teaching a language; cf. *On the Teacher* 10.29); nor did Augustine, who knew his grammar, fail to realize that language contains much more than names. What is true is that for Augustine all words signify. Words were in fact conventionally instituted as "given signs" (as opposed to "natural signs," which are not intended to signify) among human beings for the purpose of communication, teaching, and reminding: "A sign is what shows both itself to the senses and something beyond itself to the mind" (*On Dialectic* 5.7). Augustine is not consistent and clear, however, on what words signify: Sometimes he suggests that they signify incorporeal states of mind (thoughts, volitions, impulses; for example "if" signifies the speaker's doubt); sometimes he claims that words signify things, either perceptible or intelligible (cf. Kirwan 1994: 193). Although communication is the *raison d'être* of language, some form of inner language is the medium of all our thought. Even before being expressed in a natural language, what is conceived in our souls is already a "silent word," but "in no language" (cf. *Lectures on the Gospel of John* 3.14.7; *On the Trinity* 15.10.19). Augustine's original take on Plato's influential idea, also criticized by Wittgenstein (*Philosophical Investigations* 1.32), discloses the possibility of the existence of mental contents which are not (fully) communicable in any language (cf. Rist 1997).

Related Topics

7.2 Medieval Philosophy of Language.

References

General

Everson, S. (ed.) (1994) *Language. Companions to Ancient Thought 3*, Cambridge: Cambridge University Press.
Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Clarendon Press.
Modrak, D. K. W. (2006) "Philosophy of Language," in M. L. Gill and P. Pellegrin (eds.) *A Companion to Ancient Philosophy*, Oxford: Blackwell, 640–63.
Mortley, R. (1986) *From Word to Silence*, 2 vols., Bonn: Hanstein.
Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsehood*, Amsterdam: North Holland Publishing.

The Sophists

Guthrie, W. K. C. (1971) *The Sophists*, Cambridge: Cambridge University Press, 176–225.
Kerferd, G. B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press, 68–77.

Plato

Crivelli, P. (2008) "Plato's Philosophy of Language," in G. Fine (ed.) *The Oxford Handbook of Plato*, Oxford: Oxford University Press, 217–42.
Denyer, N. (1991) *Language, Thought and Falsehood in Ancient Greek Philosophy*, London and New York: Routledge.
Sedley, D. N. (2003) *Plato's* Cratylus, Cambridge: Cambridge University Press.
—— (2006) "Plato on Language," in H. H. Benson (ed.) *A Companion to Plato*, Oxford: Blackwell, 214–27.
Williams, B. (1981) "Cratylus' Theory of Names and its Refutation," in M. Schofield and M. Nussbaum (eds.) *Language and Logos: Studies in Ancient Greek Philosophy*, Cambridge: Cambridge University Press, 83–94.

Aristotle

Charles, D. (1994) "Aristotle on Names and their Signification," in Everson 1994, 37–73.
Irwin, T. H. (1981) "Aristotle's Conception of Signification," in M. Schofield and M. Nussbaum (eds.) *Language and Logos: Studies in Ancient Greek Philosophy*, Cambridge: Cambridge University Press, 241–66.
Modrak, D. K. W. (2001) *Aristotle's Theory of Language and Meaning*, Cambridge: Cambridge University Press.
Shields, C. (1999) *Order in Multiplicity: Homonymy in the Philosophy of Aristotle*, Oxford: Clarendon Press.
Whitaker, C. W. A. (1996) *Aristotle's* De Interpretatione*: Contradiction and Dialectic*, Oxford: Clarendon Press.

Hellenistic Philosophy

Barnes, J. (1999) "Language: Meaning," in K. Algra, J. Barnes, J. Mansfeld and M. Schofield (eds.) *The Cambridge History of Hellenistic Philosophy*, Cambridge: Cambridge University Press, 193–213.
Frede, D. and Inwood, B. (eds.) (2005) *Language and Learning: Philosophy of Language in the Hellenistic Age*, Cambridge: Cambridge University Press.
Long, A. A. and Sedley, D. N. (1987) *The Hellenistic Philosophers*, vol. 1, Cambridge: Cambridge University Press.

Epicureans

Atherton, C. (2009) "Epicurean Philosophy of Language," in J. Warren (ed.) *The Cambridge Companion to Epicureanism*, Cambridge: Cambridge University Press, 197–215.

Brunschwig, J. (1994) "Epicurus and the Problem of Private Language," in *Papers in Hellenistic Philosophy*, Cambridge: Cambridge University Press, 21–38.

Everson, S. (1994) "Epicurus on Mind and Language," in Everson 1994, 74–108.

Sedley, D. N. (1973) "Epicurus, *On Nature* Book XXVIII," *Cronache Ercolanesi* 3: 5–83.

Verlinsky, A. (2005) "Epicurus and his Predecessors on the Origin of Language," in D. Frede and B. Inwood 2005, 56–100.

Stoics

Allen, J. (2005) "The Stoics on the Origin of Language and the Foundations of Etymology," in D. Frede and B. Inwood 2005, 14–35.

Atherton, C. (1993) *The Stoics on Ambiguity*, Cambridge: Cambridge University Press.

Barnes, J. (1984) "Medicine, Experience and Logic," in J. Barnes, J. Brunschwig, M. Burnyeat and M. Schofield (eds.) *Science and Speculation: Studies in Hellenistic Theory and Practice*, Cambridge: Cambridge University Press, 24–68.

Bobzien, S. (1999) "Logic: The Stoics," in K. Algra, J. Barnes, J. Mansfeld and M. Schofield (eds.) *The Cambridge History of Hellenistic Philosophy*, Cambridge: Cambridge University Press, 92–157.

—— (2002) "Chrysippus and the Epistemic Theory of Vagueness," *Proceedings of the Aristotelian Society* 102: 217–38.

Frede, M. (1987) "The Origins of Traditional Grammar," in *Essays in Ancient Philosophy*, Minneapolis: University of Minnesota Press, 338–59.

—— (1994) "The Stoic Notion of a *Lekton*," in Everson 1994, 109–28.

Long, A. A. (1971) "Language and Thought in Stoicism," in A. A. Long (ed.) *Problems in Stoicism*, London: Athlone Press, 75–113.

Mates, B. (1961) *Stoic Logic*, Berkeley and Los Angeles: University of California Press.

Augustine

Burnyeat, M. F. (1987) "Wittgenstein and Augustine *De Magistro*," *Proceedings of the Aristotelian Society* s. v. 61: 1–24.

Kirwan, C. (1994) "Augustine on the Nature of Speech," in Everson 1994, 188–211.

Rist, J. M. (1997) *Augustine: Ancient Thought Baptized*, Cambridge: Cambridge University Press, 23–40.

Many thanks to the Leverhulme Trust for supporting Dr Di Lascio's research through an Early Career Fellowship in the period October 2009– September 2012.

7.2
MEDIEVAL PHILOSOPHY OF LANGUAGE

Gyula Klima

Sources and Issues

Fair warning: the title of this chapter is misleading. There is no such a thing as medieval philosophy of language. At least there was no philosophical discipline recognized in medieval thought that could be justifiably identified by us as "philosophy of language." This is not to say, however, that medieval thought did not contain systematic, philosophical reflections on language as the necessary medium of human thought and communication. On the contrary, as we shall see, the medieval philosophical literature is shot through and through with such reflections (indeed, ones that have hardly been matched in sophistication ever since), but they are not to be found in treatises classifiable under the heading of a unified philosophical discipline as we conceive of it; rather, they are to be found in works in their respective medieval disciplines.

Thus, many of these reflections can be found in treatises, commentaries, and *quaestiones* [Sweeney 2008] in the well-recognized disciplines of grammar and logic, as should be obvious. However, a great deal of interesting observations relevant to the subject (rounding out the *ex professo* considerations of the above-mentioned two) can be found in works on the soul, as well as metaphysics and theology. Of course, not all of all these works is relevant to "philosophy of language" as we conceive of it.

The descriptive/normative grammars of Latin used for teaching students the elements of Latin syntax would certainly not count as anything philosophically interesting to us. However, the characteristic doctrine of the *modists* [Marmo 1999], for example, contained in works on "speculative grammar" (*grammatica speculativa*) [Ashworth 1977; Bursill-Hall 1971], attempting to provide some general philosophical underpinning for the rules of syntax contained in the descriptive/normative works, is certainly relevant to our contemporary interest in the philosophy of language. Thus, commentaries on the Roman grammarians, especially Donatus and Priscian, as well as *quaestiones* and self-contained treatises inspired by their work, are valuable sources for medieval speculation about the fundamental features any human language must have, based on the conviction that the *modi significandi* (the ways linguistic items signify), determining proper construction or well-formedness (*congruitas*), are rooted in the *modi intelligendi* (the ways we think) common to all humans, which in turn are based on the *modi essendi* of things (the ways things are). But this idea of a "universal grammar" and especially its attempted justification was controversial, and very often challenged, especially by nominalists, in the Middle Ages [Peter of Ailly 1994].

By contrast, logic, looking at language as the medium of all human thought and reasoning, *eo ipso* deals with what are supposed to be the universal conceptual structures conveyed variously by different conventional human languages. Accordingly, medieval speculations concerning the subject matter of logic, and more specific considerations of exactly how the various syntactical structures of different human languages are supposed to convey the conceptual structures they express, and, in their turn, how those conceptual structures are related to items of extramental reality, are directly relevant to our concerns in the philosophy of language. Therefore, much of the immense medieval logical literature contains valuable insights on the medieval "theory of meaning," the *theory of signification*, most of which was inspired by Aristotle's *Categories* (*Categoriae* or *Praedicamenta*) and *On Interpretation*, especially in their translations and commentaries by Boethius. On the other hand, many important considerations concerning the medieval "theory of reference," the *theory of supposition*, were prompted by Aristotle's theory of fallacies in his *Sophistical Refutations*, giving rise to a burgeoning literature on this characteristically medieval doctrine, as the most important part of the general theory of *the properties of terms* [Read 2008]. One should also mention here the characteristic genre of the treatises on *obligations*, reflecting on the rules of engaging in the sophisticated medieval art of *disputation* [Spade 2008]. Finally, one should not ignore the immense literature on universals prompted by Porphyry's *Isagoge* ('Introduction [to Aristotle's Categories]'), discussing a five-fold classification of the different sorts of universals (the so-called "predicables", *praedicabilia*: genus, species, difference, property, and accident), organized into the system of Aristotle's categories (the *praedicamenta*, which are taken to be either the most universal genera, namely: substance, quantity, quality, relation, action, passion, time, place, position, and habit (as some garment is a "habit"), or the orderings of subordinate universals into the structures traditionally called "Porphyrian Trees," dividing genera by their differences into their species, until the most specific species are reached, whose individuals only differ in accidents, but not in any essential differences) [Gracia and Newton 2010].

Indeed, "the problem of universals" in and of itself continues to be in one form or another one of the main concerns of metaphysics as well as the philosophy of language [Klima 2008]. So it is no wonder that the most refined and subtle discussions relevant to our topic revolved around this problem (or bundle of problems), despite the fact that at least by the time of John of Salisbury *everybody* took a general Aristotelian stance, accepting the Aristotelian-Boethian criticism of Plato's "naive theory" of universals. However, this general Aristotelian stance, rejecting mind-independent universal entities existing in real being on a par with, or even more perfectly than, other real entities, admitted ever more refined articulations, leading to issues that *we* would recognize as the problems of essentialism, ontological commitment, and presentism vs. eternalism.

Besides logic and metaphysics, the medieval literature on Aristotelian psychology also contains important, relevant observations that cannot be ignored from our point of view. In fact, the whole Aristotelian rejection of the "naive" Platonic theory of universals is predicated on Aristotle's theory of universal concept acquisition, his theory of *abstraction*. Furthermore, Aristotle's distinction of the three main intellectual operations (simple apprehension, judgment formation, reasoning) is the basis of the division of the subject matter of Aristotelian logic, which, in turn forms the basis of medieval ideas that we would recognize as ones concerning issues such as generativity, compositionality, and mental language, the last of which became particularly important in the late-medieval nominalist program of "ontological reduction".

Finally, theological works contribute at least three important aspects to what may be regarded as constituting what can come under the label "medieval philosophy of language": first, the problems of scriptural exegesis (which is not to say, of course, that problems of interpretation were *only* raised in the context of *scriptural* exegesis) and, in general, the problems of sensibly talking about God gave rise to much speculation about nonliteral, metaphorical and analogical uses of language. Second, considerations of what is happening in the sacraments, such as the Eucharist or Baptism, involved intriguing reflections on performative speech acts. And third, considerations of the different languages of the Biblical tradition (Hebrew and Greek), the philosophical tradition (adding Arabic) and the different religious, cultural, and social backgrounds they conveyed as well, led to a number of interesting observations in various authors concerning linguistic diversity and the role of language in cross-cultural rational argument and understanding (where we must not forget either that scholastic writers had Latin as their second language, pretty much in the way many of us have English in contemporary academia).

Given the immense diversity of sources and issues just listed, one might get the impression that there was no unified theory specifically addressing philosophical issues concerning language as a system of signs. But we should not forget about important (parts of) systematic treatises that do address issues concerning language as a part of some general semiotic reflections on the nature and classifications of signs, such as those that can be found from the Roman period (see Augustine's *De Doctrina Christiana*), through High Scholasticism (see, for example, Roger Bacon's *Compendium studii philosophiae*, as well as the numerous reflections on Aristotle's "semantic triangle," to be discussed below, in systematic treatises of logic), to very late Scholasticism, inspiring the semiotic reflections of early and later modern philosophers (see, for example, the *Tractatus de signis* of John of St. Thomas) [Meier-Oeser 2008].

As even this sketchy introductory survey should make clear, the diversity and sheer bulk of the medieval output relevant to our topic make it impossible to provide here anything like a systematic, historical account of whatever may come under the title "medieval philosophy of language." Therefore, in the following sections I will focus only on some typical medieval treatments of *some* of the salient topics mentioned in this introductory section that we would recognize as clearly belonging to our concerns in philosophy of language.

Meaning: Signification and Imposition

Although Augustine's general definition of a sign as something that besides presenting itself to the senses also presents something to the understanding [Augustine 1975: 86, 1962: 32] was very much in circulation in the Middle Ages (and had a particularly strong influence on Roger Bacon, for example), medieval considerations of linguistic meaning usually started with reflecting on Aristotle's "semantic triangle" of *words*, *concepts*, and *things* described at the beginning of his *On Interpretation* (16a3–8), and the corresponding description of 'signifying' as establishing some understanding of a thing (cf. ibid. 16b21–22). These remarks on the "semantic triangle" and signification were generally understood as asserting that conventional linguistic items, whether written or spoken, would be meaningless sounds or inscriptions without their being conventionally associated with *naturally representative* acts of our understanding, namely: *concepts*. Indeed, medieval logicians, from Boethius onward, used meaningless articulate sounds

and the corresponding inscriptions to illustrate the point: if I utter articulate sounds like *blithyri*, *biltrix*, or *baf*, whoever hears me and does not have a language in which these sounds happen to be associated with some acts of understanding, they have literally *no idea* what I am saying; indeed, since I have no idea that I intend to convey by these sounds either, I do not really say anything by uttering these sounds, just as I am not writing anything by idly doodling on a piece of paper. Furthermore, the conventionality of linguistic signs is also clear from the fact that we could make any of these meaningless utterances meaningful by simply agreeing what we mean by them, just as we managed to make the not-too-long-ago meaningless utterance 'google' to mean the popular internet search engine, and later on, by extension, the act or activity of searching on it.

Establishing the conventional relationship between utterances and acts of our understanding is what the medievals referred to as *imposition*, whether this was done (perhaps, even ceremonially) by some name-giver(s) (as happened with 'google', the name) or simply by introducing and picking up convenient usage (as probably happened with 'google', the verb).

Understanding the relations of the Aristotelian semantic triangle in this way may seem perfectly harmless and uncontroversial. But already some medieval authors were wary of what *we* might describe as the 'representationalism' apparently involved in this position. Referring to a 'magna altercatio' (great controversy) over the issue, John Duns Scotus (1265/66–1308) argued that if our words signified our concepts, then we would all speak our private languages, talking only about our concepts, and never about the things we live among, which is absurd [Pini 2001, 1999; Perler 1993, 2003].

However, Scotus may have used the phrase *magna altercatio* only to dramatic effect, to emphasize the importance of his own position. Actually no medieval thinker interpreted the Aristotelian semantic triangle as amounting to a commitment to talking only about our concepts. For whether they described the relation established by imposition between words and concepts as (immediate) *signification* or *subordination*, and whether they took our words to signify things primarily or secondarily, they all agreed on the basics outlined above that without associating articulate utterances with acts of understanding, those utterances are not significative, and that when by an act of imposition they do become significative, then they will *ultimately signify* whatever we conceive by the concepts to which they are *subordinated* or which they *immediately signify*. An utterance without being subordinated to a concept is just an articulate sound that may sound like a word of a language, but it is not; it will gain its semantic properties through its subordination to a concept, just as the inscription conventionally recording it will inherit *its* semantic properties, and both the utterance and the inscription will signify just those things that we conceive of by the concept and just in those ways in which the concept represents them.

So, our words signify whatever we conceive of by means of the corresponding concepts. But does this mean, then, that we cannot talk about our concepts, but only about the things conceived by these concepts? This does not follow, for at least two reasons: First, we can form concepts of concepts (such as the concepts to which the words 'concept', 'species', 'genus', 'negation', etc. are subordinated, as in saying 'Negation is a syncategorematic concept'), and so we can use the words subordinated to these concepts to talk about our concepts; second, we can even use words subordinated to concepts that conceive of things that are not concepts to refer to these concepts and not to the things conceived by these concepts (as when we say, 'Man is a species').

The first reason is based on our ability to form concepts of concepts, or as the medievals called them, *second intentions*. Although in the secondary literature these are some-

times referred to as "second-order concepts", the medievals' second intentions are not to be confused with what after Frege we call "second-order concepts." Frege's second-order concepts are not concepts whereby we *conceive of* other concepts; rather, they are concepts that *operate on* other concepts, that is to say, they are *syncategorematic* concepts [Klima 2006]. By contrast, second intentions are *categorematic* concepts, by which we conceive of concepts insofar as they are concepts, that is, in terms of their representative, and not in terms of their ontological, features. For instance, when I say, 'Negation is a syncategorematic concept' both the subject and the predicate of this sentence are what medieval authors called "names of second intention" or "names of second imposition", that is, names that are subordinated to concepts whereby we conceive of concepts in terms of their function. However, if I say 'The concept of negation is a quality of the human mind', the subject term of this sentence is a term of second intention, whereas the predicate is a term of first intention, describing the concept referred to by the subject in terms of its ontological features (at least on one medieval theory of concepts).

The second reason why we *can* talk about our concepts, even if our words ultimately signify the objects of our concepts, and we "normally" use our words to talk about those objects is that we *can* use our words to refer to the concepts themselves to which these words are subordinated. For instance, if I say 'Man is a species', the subject term of this sentence, insofar as it is true, is not used to refer to what it is ordinarily used refer to, namely, individual human beings (as in 'Every man is an animal'), but rather to the specific concept of individual humans (whereby these individuals are conceived precisely insofar as they are human), for that concept is a species on the medieval interpretation of 'species' (in line with Porphyry's definition, according to which a species is something predicable of several numerically distinct individuals). Note that in this case, in contrast to the former, we do not refer to a concept by means of a term subordinated to a second intention, a concept of a concept. In this case we use the term subordinated to a concept to refer to this concept; so, it is not a term of second intention or second imposition that is used to refer to a concept (possibly other than the concept to which this term is subordinated) but a term of first imposition that is used with the meaning (signification) it ordinarily has (i.e., subordinated to the concept that gives its signification) to refer to the concept to which it is subordinated.

To be sure, since an act of imposition is voluntary and hence subordination is conventional, the same utterance (or the corresponding inscription) may be used not only with the same signification whenever it is used but in different senses, imposed on different concepts. Thus, the same term may be used according to totally different and unrelated significations, as imposed on totally different concepts, leading to *equivocation* (as in 'I am going to the bank' meaning either the river bank or the financial institution), or on essentially different concepts that evoke similar imagery or some meaningful associations in the hearer's mind (as in 'The meadow is smiling', meaning that it is flowering), leading to *metaphor*, or on essentially related concepts, one of which is the result of some analogical extension of the other (as in 'I see your point', meaning that I understand it), leading to *analogy*. In fact, relativizing signification to imposition, and making imposition dependent on the intention of the speaker or interpreter, provided medieval scholastics with a very flexible framework for handling all sorts of improper uses and meanings, including irony, sarcasm, humor, or even cases when we use a term as imposed on a concept of others not shared by us (as when a Christian talks about the 'gods' of the pagans).

However, the foregoing example of 'Man is a species' illustrated not any of these

improper uses, affecting what is meant by the term (its signification), but the case when the same term, according to the same imposition, i.e., with the same meaning it normally has, can be used to stand for something other than what it is "normally" used to stand for. Therefore, since the same term with the same meaning or signification can be used to refer to different things, *meaning* must be distinguished from *reference*, or as our medieval colleagues would put it, *signification* must be distinguished from *supposition*. Thus, on top of the variability of meaning resulting from the variability of imposition, the medieval theory of supposition adds another layer of variability resulting from actual usage in context, insofar as the same term used with the same signification can still be used to stand for different things, or, as we shall see, even for the same things, but in different ways.

Reference: Supposition, Appellation, Ampliation and Restriction

Supposition is a property of terms in the context of a proposition, namely: their property of standing for a thing or several things that the proposition is about. On the basis of this general characterization, one might think that a medieval theory of supposition is just the same as a modern theory of reference. However, even on this very general level, there are some fundamental differences. (1) Medieval authors in general insisted that supposition is a property of terms in a propositional context, that is to say, supposition is essentially context-dependent: It simply makes no sense to ask what the supposition of the term 'man' is; although it does make sense to ask what the supposition of the term 'man' is in the proposition 'Every man is an animal' or in the proposition 'Every man is a noun', and we would get different answers in these two cases. (2) Furthermore, *some* medieval authors, most notably Peter of Spain, would assign different types of supposition to the same term in the same proposition, depending on whether it is the subject or the predicate of that proposition, while most authors would assign it the same *type* of supposition but possibly in a different *mode*. (3) Again, supposition may only be just one of the several referential functions a term may have in a proposition whereby it may contribute to the compositional semantics of the proposition in which it occurs. On some medieval theories—most importantly, on Buridan's theory of *appellation*, the *oblique reference* to the *connotata* or even to the *immediate significata* of some terms—their *appellation* is just as important in determining the truth or falsity of propositions as their *supposition* is, since their supposition is in fact determined in part by their appellation. (4) Finally, supposition may be *restricted* or *ampliated* by the propositional context of a term; indeed, it may be ampliated even to non-existents but still without any "ontological commitment" to some weird nonexistent quasi-entities [Klima 2009: 162–174]. To clarify these summary remarks, we should consider them one by one in their turn.

1. If I say 'Socrates is a man' and if I say 'Socrates is a noun' the intended referent once is the man Socrates, once the noun we use to refer to the man in the other sentence (or indeed *that* noun in the other sentence and/or any other token noun of the same type). Of course, in contemporary written English we would indicate our intended reference to the noun by the use of quotes, but in spoken English we can at most gesture (and/or say, 'quote/unquote') to indicate quotes, and in the primarily oral culture of Medieval Latin there was no such device available at all. Instead, medieval logicians distinguished *personal* and *material* supposition to distinguish the cases when a term is taken to stand for what it ultimately signifies

(according to nominalists—or what actually instantiates what it ultimately signifies, according to realists; more on this later) from the case when it is taken to stand for itself and/or for any other token of the same type. (Late medieval logicians did introduce, however, a device similar to modern quotes: They used the prefix 'li' or 'ly' in front of a term to indicate that it is meant to stand in material supposition; some even used a slash in the place of an unquote, especially when they used a complex term or a proposition in material supposition.) Besides these two types of supposition, most medieval logicians also distinguished *simple* supposition, when a common term is taken to stand for the concept it immediately signifies, whether the concept in question was taken to be some individualized mental act *by which* its object is conceived or *that which* is conceived, *precisely as conceived* by such a mental act. For instance, the nominalist William Ockham in his mature theory held that in the proposition 'Man is a species' the subject term stands for individualized mental acts of individual human minds whereby they conceive of all humans, past, present, future, or merely possible, indifferently, in a universal manner. In his earlier theory, however, he held that the concept for which this term in this proposition stands is a universal object-of-the-mind, a so-called *fictum*, which (as its name indicates) is not a real being, for its being consists in nothing but its-being-conceived. So-called "realists," however, such as Burley or Scotus, or even the more "moderate" realist Aquinas would assign some more mind-independence to these objects-as-conceived as far as their "foundation in reality" is concerned, although in their ontologies they would still refrain from treating them on a par with really existing singulars. But all these are just rather subtle *metaphysical* differences among these authors; semantically they would all by and large agree that a term in simple supposition stands *not* for the particulars that fall under it, but rather for its immediate significate (that is, the concept to which it is subordinated), whatever the ontological status of that immediate significate is [Burley 2000; Klima 1993].

2. One of the most influential medieval logicians, Peter of Spain, famously argued that the predicate terms of categorical propositions always have to stand in simple supposition, while the same terms as subjects can have any other kind of supposition. Accordingly, on his position, in 'Every man is a man' the subject term would have personal supposition, standing for individual humans, the individuals informed by the ultimate significata of the term, namely, by individualized human natures, but the predicate would have simple supposition, standing for its immediate significate, namely, human nature as conceived in abstraction from its individuating conditions. Peter's position was actually quite unique and going against "the mainstream" conception which would have accorded personal supposition even to the predicate term of this sentence, although in a *mode* different from the subject, characterized in terms of a different type of "descent to the singulars." For when a common term stands for the things that fall under it (whether they are its ultimate significata or the things informed by its ultimate significata, on a nominalist or a realist conception, respectively), according to the medieval theory of common personal supposition, the term can stand for these *same things in different ways*, depending, again, on the propositional context. For instance, in 'Every man is a man', the subject term would by all authors be regarded as having *confused and distributive* personal supposition, characterizable by means of a "descent to the singulars," that is, an equivalent *conjunctive proposition*, in which the common subject term of the original is replaced in each member of the conjunction by a singular term referring to one of the sup-

posita of the original term, as in 'This man is a man *and* that man is a man *and* . . .' and so on, for all humans. By contrast, in 'Some man is a man' a similar descent to the singulars under the subject term is only possible by means of a *disjunctive proposition*, as in 'This man is a man *or* that man is a man *or* . . .', whence this term was said to have *determinate* supposition. According to "the common doctrine," the predicate term of a universal affirmative proposition could have neither of the two above-mentioned modes, for no equivalent descent under this term is possible, whether by means of a conjunctive or by means of a disjunctive proposition. (Clearly, 'Every man is this man or every man is that man . . .' as well as 'Every man is this man and every man is that man . . .' is false, provided there is more than one man, while the original 'Every man is a man' is true.) However, an equivalent descent by means of a *disjunctive term* is still possible: 'Every man is this man or that man or . . .' is true just as well as the original is in all possible cases in which the original is true. (Concerning these equivalences one also has to take into account medieval considerations concerning the existential import of universal affirmatives, to be discussed in the next section.) This case, therefore, was regarded as providing an instance of a different mode of common personal supposition, commonly called *merely confused* supposition. Considerations of completeness and symmetry would also seem to require a fourth mode, namely: one in which descent is made to a *conjunctive term* (for instance, descending from 'Only all men are running' to 'Only this man and that man and that man . . . are running', and some late medieval authors did in fact discuss this mode under the name of *conjunct supposition* (the so-called *suppositio copulatim*) [Read 1991a, 1991b]. In fact, the resulting system of descents and corresponding ascents can be quite naturally expanded into a powerful analytic tool, matching and even surpassing the analytic abilities of standard quantification theory (the different types of descents distinguishing universal and existential quantifiers with different relative scopes), applicable even to recently often discussed cases of doubly numerically quantified sentences, such as 'Two examiners marked six scripts' [Klima and Sandu 1990]. Nevertheless, most medieval authors were quite content with only the three "standard" modes, apparently not motivated by the concerns we have about such sentences. In any case, returning to Peter's position on the predicate term of universal affirmatives, he first remarks that assigning simple as well as merely confused personal supposition to such a term would not be inconsistent, for these two modes are not incompatible. The term might exercise the function of referring to human nature in general, conceived in abstraction from the individuals, as well as the function of referring to the individuals informed by this nature, not determinately or distributively, but merely confusedly, as by means of the corresponding nominal disjunction. However, he next argues that a common predicate term simply does not have this function. His argument can be briefly restated as follows. In 'Every man is an animal' a genus is predicated of one of its species. But the nature of the genus *multiplied* in the supposita of the species *is not* the genus (which is a universal, that is, something that is *one* over many, and thus *not multiplied* with the multiplicity of the many things participating in it). Therefore, it is not the nature of the genus multiplied in the supposita of the species that is predicated here. But then, since the predicate of this sentence stands for what is predicated, which is not the nature of the genus multiplied in the supposita of the species, the predicate cannot have confused supposition, which would require this multiplication. Based on this conclusion, Peter states that although from the

point of view of logic the nature signified by 'man' in its supposita is one, in the nature of things each man has his own humanity and these humanities are distinct on account of the matter they inform. Likewise, the nature signified by the term 'animal' in individual humans is one from the point of view of logic (*secundum viam logice*) but is multiplied in these individuals in the nature of things (*secundum viam nature*). So, the multiplication of animalities has nothing to do with the semantic function of the predicate of 'Every man is an animal'; indeed, in the nature of things we find the same multiplication of animalities (namely: the multiplication of individualized animalities of human beings), even when we consider 'Every man is white' or 'Every man is black', where there is no mention of these animalities at all.

3. In accordance with the foregoing, despite its differences from modern conceptions of reference, supposition can still be regarded as a term's context-dependent reference, insofar as supposition is the context-dependent feature of the term determining *what* and *how* it stands for in a proposition in which it is used. But direct reference to what a term stands for is not the only possible referential function a term can have, for it may also obliquely refer to other things in relation to the things it stands for. The most comprehensive medieval account of this sort of co-reference is John Buridan's theory of *appellation*. On Buridan's interpretation, *connotative* (or *appellative*) terms signify the things that fall under them not absolutely as *absolute* terms do, but in relation to other things, which are their *connotata*. Therefore, in a proposition in which they stand in personal supposition, besides suppositing for the things they signify, such terms also obliquely refer to or *appellate* the things they connote. For instance, the term 'father' is a connotative term, for it signifies men not absolutely, but in relation to their children (in fact, it signifies all men, past, present, future or merely possible, indifferently, just as the term 'man' does, but in relation to their past, present, future, or merely possible children), who are this term's *connotata*. Accordingly, in the context of a present tense proposition, as in 'Every father loves his child or children', the subject term will supposit for those of its *significata* that are actually related to its connotata in the way the signification of the term demands, that is to say, it will only supposit for those men who presently have children. Thus, clearly, *appellation*, the oblique reference to the *connotata* of a connotative term, has an important role in determining the direct reference, the *supposition* of the same term, and hence the truth of the proposition in which it occurs. For, given the fact that the subject term of this proposition stands in common, personal, confused and distributive supposition, it is true only if the conjunction 'This father loves his child or children and that father loves his child or children and . . .' is true. However, if in a possible situation no man actually has a child, than no man can be the *suppositum* of 'this father', whence no man can be a suppositum of 'father'. Accordingly, when the subject term of a universal affirmative supposits for nothing (possibly on account of the lack of the relevant *appellata*, as in this case), then the proposition is false; that is to say, it must have *existential import*. But it is not only appellative terms that can have such oblique reference: For absolute terms occurring in an oblique case in a complex term also appellate their ultimate significata, and according to Buridan *all terms* appellate their immediate significata (the concepts to which they are subordinated) as the direct objects of *intentional* verbs. For example, in 'Every donkey of some man is running' (*Omnis asinus alicuius hominis currit*), the complex subject term supposits for donkeys, but

only for those that belong to some man or another, on account of the appellation of men by the term in the oblique case (*hominis*). In this case, the term clearly appellates its ultimate significata. By contrast, in 'I know the one approaching' (*Cognosco venientem*) or 'I owe you a horse' (*Debeo tibi equum*), Buridan explains the peculiar behavior of the direct object of the verbs (lack of existential import, breakdown of the substitutivity of identicals) by claiming that these terms in these contexts appellate the concepts to which they are subordinated, for the mental acts signified by these verbs (as opposed to bodily acts) concern their objects only insofar as they are cognized precisely by means of the concept to which the direct object of the verb is subordinated.

4. In fact, Buridan argues that as a result of this sort of appellation of concepts (*appellatio rationis*) the *supposition* of the terms construed with intentional verbs is *ampliated*, i.e., extended beyond the range of presently existing things even in a present tense context, for by means of the concepts appellated by these common terms we conceive of all things that fall under it, past, present, future and merely possible indifferently, in a universal manner. But this sort of *ampliation* of supposition is achieved not only by intentional verbs but also by tenses and modalities—in general, whatever we would recognize as *intensional* contexts. These contexts, therefore, would seem to force what *we* would recognize as "quantifying over nonexistents." However, for medieval logicians, given the variability of supposition, there is not one fixed domain of quantification in which we are ontologically committed to every item we quantify over: We are only committed to the existence of those things of which it is true to affirm in a present tense proposition that they exist. Of course, in the appropriate intensional contexts we can make reference to things that existed in the past, or will exist in the future, or can exist in a possible situation, given our awareness of such items through our relevant mental acts, but reference to these items in ampliative contexts will never entail a proposition about their actual existence, so they are not parts of our ontology at all. This is not to say, however, that medieval realists would have refused to admit in their ontologies items that medieval nominalists would find to be "weird entities," which, therefore, they "analyzed away" by working out the details of how our "mental language" can map our category-rich spoken and written languages onto a parsimonious ontology, containing only two or three distinct categories of entities.

Therefore, after these points, which served to illustrate the main differences between medieval and modern theories of reference and quantification in general, we should turn in conclusion to some of the fundamental differences between medieval nominalist and realist strategies in constructing their logical-linguistic theories, in particular with reference to their rather different conceptions of compositionality and mental language.

Compositionality, Mental Language, and Ontological Commitment in Medieval Realism and Nominalism

As the foregoing discussion of Buridan's theory of appellation illustrated, medieval logicians were very much aware of the compositional character of complex expressions. Indeed, a late-15th-century logician, Hieronymus Pardo, explained linguistic generativity precisely in terms of compositionality. He argued that in order for a proposition

to signify what it does, it does not have to be imposed to signify separately as a whole, for the signification of the whole results from the significations of the parts, and this is why we are capable of generating a potential infinity of meaningful propositions [Pardus 1505: 1].

However, since in accordance with the generally accepted theory of imposition and meaning, items in our spoken and written languages inherit their semantic features from our concepts, the naturally representative units of human thought, compositionality was also regarded as primarily belonging to our natural mental representational system, our *mental language* [Panaccio 1999, 2004; Klima 2009: 37–120]. This medieval "language of thought hypothesis" was not so much regarded as a "hypothesis", but as the natural, most plausible explanation of the facts of how any human language works. Accordingly, the syntactical constructions of natural languages were regarded as exhibiting, but sometimes as masking, the underlying conceptual constructions, which therefore could be explicated by using the syntactical constructions exhibiting them to analyze or expound the constructions masking them.

Accordingly, nominalist logicians, such as Ockham, Buridan, and their ilk, argued that the syntactic simplicity of some connotative terms may actually mask a complex conceptual structure on which the simple connotative term is merely imposed for convenience, but the actual complex concept to which such a term is subordinated can be explicated by means of the nominal definition of the term, a complex expression whose syntactical structure matches the compositional structure of the complex concept in question. In fact, this type of analysis provided the most powerful conceptual tool for nominalist philosophers to eliminate what they regarded as the absurd or just unnecessary ontological commitments of their realist opponents. Take for instance a privative term (one that indicates a defect, the lack of a naturally desirable characteristic), such as 'blind'. For realist philosophers (even for such a "moderate" realist as Aquinas), this term, and primarily the concept to which it is subordinated, has to signify the lack of sight in animal that ought to have it by nature. But what is such a *lack* of sight, a *privation*? It certainly cannot be an entity on a par with a sight, which was taken to be a positive quality, a natural power enabling an animal to see, for blindness is precisely the lack, the nonbeing of such an entity. But on the prevailing theory of predication of the realists, an affirmative predication is true just in case the ultimate significate of the predicate is actually present in the suppositum of the subject; thus, 'Homer is blind' is true, just in case the blindness of Homer is present in Homer's eyes. But since the presence or existence of the blindness is nothing but the absence or nonexistence of his sight, Homer's blindness cannot be regarded as an entity existing in the same sense as his sight would, so it was classified as a special kind of entity, a so-called being of reason (*ens rationis*), the presence of which is conditioned both on how things are (in this case, on Homer's not having sight) and on how they are conceived, namely: in terms of the concept of blindness whereby we conceive of this situation by forming this privative concept.

By contrast, nominalists, unwilling to buy into the dubious ontology of such quasi-entities, argued that the nominal definition of 'blind', namely 'animal not having sight' clearly indicates that there is no need to assume the ("halfway") existence of such spurious entities, for this nominal definition shows that the concept to which 'blind' is subordinated consists of two absolute concepts, whereby we merely conceive of "ordinary" entities, namely, animals and sights: absolutely, a connotative, relative concept, whereby we conceive of these same entities in relation to one another, and

the syncategorematic concept of negation, which simply has the effect that the resulting complex concept, signifying animals and connoting their sights, will supposit in a mental sentence precisely for those animals that do *not* have their sight. Thus, on this analysis there is no need to assume the existence of any spurious entity, only ordinary substances and their qualities.

There is, however, a curious thing about the compositionality of mental language. Medieval philosophers were generally committed to the view that our mental acts must be ontologically simple qualities of our minds. How could they possibly talk, then, about complex concepts, if they are supposed to be simple entities? And how could the semantic values of these complex concepts be dependent on the semantic values of their components, if they are simple, meaning that they cannot have any parts? The answer to these questions is that semantic compositionality in mental language does not have to go together with syntactic construction in the way it does in spoken and written languages. For a complex concept does not have to be complex on account of being some (structured) aggregate of several distinct parts; it is enough if it is merely *semantically complex* in the sense that its semantic values are functionally dependent on the semantic values of other (simpler or simple) concepts, and possibly on other contextual factors, such as imposition, time, place, speaker, listener, pointing, etc. [Klima 2001: xxxvii–xxxix]. In fact, quite interestingly, the "realist" Cajetan argues for the same idea of compositionality in mental language, as not requiring any sort of ontological aggregation of several simple concepts: The semantically complex concept can be ontologically as simple as the semantically simple ones, for the semantic complexity of the complex concept merely consists in the functional dependency of its semantic content on the semantic content of the semantically simple ones.

What nominalists and realists do disagree on is rather what this semantic content is, and, accordingly, how one should conceive of the task of the logician in accounting for it in a compositional semantics. Very briefly, realists are "ontologically nonchalant" in their semantics, and start worrying about ontological commitments only in their metaphysics. By contrast, nominalists are "ontologically parsimonious," eliminating any unwanted or unneeded ontological commitments in their semantics, although they may actually allow certain types of entities in their metaphysics or physics for other than semantic reasons, for as far as their semantics is concerned, they could do just as well without them. But the reason for this difference in strategies is exactly how they conceive of *semantic content*. For realists, the semantic content of our mental acts is precisely the *information* they carry about their ultimate objects, no matter what sorts of concepts we are talking about and what sorts of objects these are. So, for realists all our concepts are representative, whether they are categorematic or syncategorematic, simple or complex. This is how they end up in their ontologies with entities and quasi-entities corresponding to all sorts of linguistic categories—such as modes-of-things-as-conceived, corresponding to syncategoremata, universals, privations, negations, relations of reason, all *beings of reason*—corresponding to categorematic terms involving in the formation of their content some mental operation; and *enuntiabilia*, *dicta*, *complexe significabilia* or *real propositions*, corresponding to mental propositions or judgments. By contrast, for nominalists, the contents of our mental acts are the things conceived themselves. Accordingly, for a staunch nominalist like Buridan, the semantic contents of our propositions will not be anything distinct from the contents of their terms: both our terms and our propositions are *about the same things* but representing them *differently*—singularly or universally, affirmatively or negatively, in simple or a complex fashion—whereas all

these differences in the mode of representation are accounted for in terms of conceptual construction or other *genealogical* features of our concepts, determined by the process of their formation to be accounted for in *psychology*. Thus, for Buridan, logic will have nothing to do with semantic content in the way realists conceived of it: The truth or falsity of our propositions will depend only on the identity or nonidentity of the suppositа of their terms, whereas the information conveyed by them will be a matter of psychology. In a word, nominalist logic is *extensionalist*, as opposed to the essentially *intensional* logic of the realists.

Bibliography

Ashworth, E. J. (1977) *Tradition of Medieval Logic and Speculative Grammar*, Toronto: Pontifical Institute of Mediaeval Studies.

Augustine, *De dialectica*, ed. Jan Pinborg, Boston: Dordrecht, 1975

Augustine, *De doctrina Christiana*, ed. Joseph Martin, Turnhoult: Brepols, 1962.

Burley, W. (2000) *On the Purity of the Art of Logic: The Shorter and Longer Treatises*, translated and edited by P.V. Spade, New Haven and London: Yale University Press, 86–90.

Bursill-Hall, G. L. (1971) *Speculative Grammars of the Middle Ages: The Doctrine of the* partes orationis *of the Modistae*, Approaches to Semantics, 11, The Hague: Mouton.

Dominik, P. (2003) "Duns Scotus's philosophy of language," in: T. Williams (ed.) *The Cambridge Companion to Duns Scotus*, Cambridge: Cambridge University Press,161–192.

Gracia, J. and Newton, L. (2010) "Medieval Theories of the Categories," *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), forthcoming http://plato.stanford.edu/archives/sum2010/ entries/medieval-categories/

Klima, G. (1993) "The Changing Role of *Entia Rationis* in Medieval Philosophy: A Comparative Study with a Reconstruction," *Synthese* 96: 25–59.

Klima, G. (2001) *Introduction* to John Buridan: *Summulae de Dialectica*, an annotated translation with a philosophical introduction, New Haven: Yale University Press.

Klima, G. (2006) "Syncategoremata," in: Keith Brown (ed.) *Elsevier's Encyclopedia of Language and Linguistics*, 2nd Edition, Oxford: Elsevier, 353–356.

Klima, G. (2008) "The Medieval Problem of Universals," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*, http://plato.stanford.edu/archives/win2008/entries/universals-medieval/

Klima, G.and Sandu, G. (1990) "Numerical Quantifiers in Game-Theoretical Semantics," *Theoria* 56: 173–192.

Marmo, C. (1999) "The Semantics of the *Modistae*," in Sten Ebbesen and Russell L. Friedman (eds.), *Medieval Analyses in Language and Cognition*, Acts of the Symposium, 'The Copenhagen School of Medieval Philosophy', January 10–13, 1996, Copenhagen: Royal Danish Academy of Sciences and Letters and C. A. Reitzels Forlag, 83–104.

Meier-Oeser, S. (2008) "Medieval Semiotics," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, http://plato.stanford.edu/archives/fall2008/entries/semiotics-medieval/

Panaccio, C. (1999) *Le discours intérieur de Platon à Guillaume d'Ockham*. Paris: Éditions du Seuil.

Panaccio, C. (2004) *Ockham on Concepts*. Aldershot-Burlington: Ashgate.

Pardus, H. (1505) *Medulla Dyalectices*, Anabat-Gerlier: Paris.

Perler, D. (1993) "Duns Scotus on Signification," *Medieval Philosophy and Theology* 3: 97–120.

Perler, D. (2003) "Duns Scotus's Philosophy of Language", in T. Williams (ed.), *The Cambridge Companion to Duns Scotus*. Cambridge: Cambridge University Press, 161–192.

Peter of Ailly (1994) *Destructiones modorum significandi*, edited by Ludger Kaczmarek, Amsterdam: John Benjamins Publishing Company.

Pini, G. (1999) "Species, Concept, and Thing: Theories of Signification in the Second Half of the Thirteenth Century," *Medieval Philosophy and Theology* 8, 2: 21–52.

Pini, G. (2001) "Signification of Names in Duns Scotus and Some of His Contemporaries," *Vivarium* 39: 20–51.

Read, S. (2008) "Medieval Theories: Properties of Terms," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, http://plato.stanford.edu/archives/fall2008/entries/medieval-terms/

Read, S.L. (1991a) "Descensus copulatim: Albert of Saxony vs. Thomas Maulfelt," in J. Biard (ed.), *Itineraires d'Albert de Saxe: Paris-Vienne au XIVe*, Paris: J. Vrin, 71–85.
Read, S.L. (1991b) "Thomas of Cleves and Collective Supposition," *Vivarium* 29: 50–84.
Spade, P.V. (2008) "Medieval Theories of *Obligationes*," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, http://plato.stanford.edu/archives/fall2008/entries/obligationes/
Sweeney, E. (2008) "Literary Forms of Medieval Philosophy," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, http://plato.stanford.edu/archives/fall2008/entries/medieval-literary/

7.3
MODERN PHILOSOPHY OF LANGUAGE

Michael Losonsky

1 Introduction

The history of modern philosophy roughly covers seventeenth- and eighteenth-century European philosophy, a prolific period spanning from Descartes to Kant. It is in this period that the philosophy of language first becomes a viable subdiscipline, as the greatest philosophers of the time turned to the study of language in a manner very different from their forebears of previous centuries. In the modern period, human natural language became a distinct object of systematic philosophical study expressed in substantial and self-contained texts. The semantic, grammatical, and pragmatic structures of natural language come to be seen as domains worthy of the kind of sustained study that before had been devoted to logic or metaphysics. Moreover, philosophers of this era turned to language in order to solve problems in other areas of philosophy. The study of language was not just a side interest or a source of evidence for already established philosophical principles but was seen as necessary for solving philosophical problems. Finally, central themes and strategies that define the philosophy of language in later centuries, including contemporary analytic philosophy of language, emerge during this period.

Many factors contributed to the turn to language in the seventeenth century, a major one of which was a loss of confidence in the received Aristotelian view about the relations between language, mind, and reality. Aristotle had maintained that the spoken sounds of natural languages are "signs of affections of the soul" and "these affections are likenesses of actual things" (*De Interpretatione* 16a4–7). Moreover, while there is a diversity of natural languages, affections of the soul as well as actual things are "the same for all." All natural languages mirror a shared set of ideas and concepts that in turn mirror a shared reality.

Toward the end of the sixteenth-century, philosophers worried that ideas are not the same for all and that they do not reliably mirror reality. The scientific revolution had cast doubt on the assumption that human ideas represent actual things. Moreover, philosophers had come to fear that natural languages were inherently vague or ambiguous and that they contain expressions for outdated and inaccurate conceptions of reality that could impede scientific progress. This loss of confidence that there is a deeper unity below the surface of linguistic diversity was a key factor in the emergence of the philosophy of language.

There were two basic responses to the ruptures in the Aristotelian unity of mind, language and reality. Roughly speaking, linguistic universalists sought to demonstrate that

there indeed was a deeper unity, either a universal grammar or a shared logical structure, beneath the diversity of natural languages. Linguistic pluralists viewed the diversity of linguistic structures as an essential feature of language. There is no deeper unity to all languages because their diversity reflects the diversity of human minds. What follows is an overview of both types of response.

2 Linguistic Diversity

Francis Bacon (1561–1626)

Bacon does not devote a major text to natural language, but his influential writings raised issues that turned seventeenth-century philosophy to language. Bacon's lifelong project was to single-handedly replace Aristotle's logic. He called his logic the *New Organon* (1620) to indicate that he intended it as a replacement of Aristotle's treatises on logic, which had been collected by Aristotle's students under the heading *Organon* (The Instrument). Bacon rejected Aristotle's focus on the deductive syllogism, believing that it only serves the justification of what is already known, not the discovery of new knowledge. He held that logic should give "*True Directions for the Interpretation of Nature*," (the subtitle of the *New Organon)* and Bacon argued that the proper instrument for discovery is what he called "induction," that is, a method for systematically organizing our observations and drawing conclusions from them.

Bacon believed that his method was a remedy for certain systemic errors in human cognition that he described as the "idols" or illusions of the mind. Bacon identified human natural languages as a source of illusions, namely the idols of the marketplace or town square [*idola fori*]. Language is used to communicate human knowledge, including scientific knowledge, but the terms of natural languages are defined according to the understanding "of the vulgar" or the uneducated. Consequently, many words do not refer to anything and many words that do refer are ambiguous or poorly defined.

Introducing better definitions informed by science cannot stop this illusion. Definitions, according to Bacon, are always verbal definitions that depend on other words, and inevitably some of these more basic words have meanings and uses determined by uneducated people because language must "suit the understanding of the common people" (2000: 42). Ordinary appearances, however, rarely match how things really are. Only observation directed and organized by a scientific method gives us a better insight into nature, and only an educated elite is capable of scientific work.

For Bacon, then, human language is not just a mirror of the human mind, but it also *structures* the mind. While human beings assume that "their reason governs words," in fact "it is also true that words retort and turn their force back upon the understanding" (2000: 48). Bacon has a striking metaphor for this feedback loop: "Words, as a Tartar's bow, do shoot back upon the understanding of the wisest, and mightily entangle and pervert judgment" (2001: 137–8).

Since the harmful effects of language on human cognition cannot be avoided, Bacon is reserved about attempts to construct a scientific or "philosophical" language consisting only of words that "are the footsteps and prints of reason." The congruence between "words and reason is handled *sparsism*, brokenly" (2001: 141–2). Nevertheless, Bacon is not entirely dismissive of the possibility of a philosophical language, and he licensed such projects. While there are only a scattered few words that are the footsteps of reason, words are not entirely deficient in this respect and hence Bacon says he

"cannot report it deficient" and recommends that there should be a science devoted to the development of a philosophical language. However, this language would consist only of conventional and arbitrary signs imposed voluntarily by human beings. Bacon is wholly dismissive of the idea that a philosophical language can be constructed out of "real characters," such as ideograms or hieroglyphs, where there is a likeness or similarity between the character and what it represents.

Thomas Hobbes (1588–1629)

Hobbes, Bacon's personal secretary between 1619 and 1623, accepted Bacon's critique of natural language. Hobbes famously wrote, "words are wise mens counters, they do but reckon with them; but they are the mony of fooles" (1996: 28–9). Only a fool uses words without paying attention to their meanings, namely the ideas they signify. The "first abuse" of speech is speaking with wrong or no definitions, and correct speech pays attention to the meanings of words and supplies univocal, clear, and consistent meanings. Accordingly, the first step of any science is to start with "the right Definition of Names" (1996: 28).

However, unlike Bacon, Hobbes assigns a positive role to language in human cognition. Natural language is so important for Hobbes that he devotes a full chapter of the *Leviathan* to natural language just before the fifth chapter "Of Reason and Science" (1996: 20). He believes that without language human beings cannot generalize, and without generalization neither reasoning nor science are possible. For example, without language human beings cannot think about classes, properties, or kinds. Hobbes maintains that there are only individual things and there is "nothing in the world Universall but Names; for the things named, are every one of them Individual and Singular" (1996: 26). Moreover, language allows human beings to order their thoughts over longer periods of time and order them with a method involving general principles. For example, arithmetic calculations as well as logical deductions require the languages of arithmetic and logic. Accordingly, in the *Leviathan* Hobbes defines reason as "nothing but the *Reckoning* (that is, Adding and Subtracting) of the consequences of generall names agreed upon, for the *marking* and *signifying* of our thoughts" (1996: 32). In other words, reasoning is a kind of symbolic computation, and for this reason Hobbes has been seen as the grandfather of Artificial Intelligence (Haugeland 1989). Hobbes even tries to sketch the rudiments of a language for computing logical inferences on the basis of a finite set of operations (Hobbes 1981).

Since language is essential to reason and science, creatures without language, including young children that have not yet acquired language "are not endued with Reason at all." Moreover, reason itself is not innate to human beings but acquired through "study and industry" because for Hobbes languages are artifacts constructed by human beings (1996: 21–3). Thus, all that depends on language, including science, has a conventional component. Hobbes believed that science could achieve only "conditionall Knowledge," namely knowledge of the consequences of the conventional definitions that stand at the beginning of any science.

Since all languages are products of voluntary activity, there will be a diversity of languages answering to the diversity of psychological and physical constitutions. Bacon already described human psychological diversity as "Idols of the Cave," namely errors due to "the peculiar constitution, mental or bodily, of each individual; and also education, habit, and accident" (2000: 41). For Hobbes it is a fundamental fact of the human

condition that while nature is the same for all people, their individual natures, dispositions, desires, and needs are not the same, and so there is a "diversity of our reception" of nature (1996: 31, also 25).

Accordingly, Hobbes's philosophy of language has no room for a philosophical language or a universal language other than a language that for political reasons has been imposed on people by a sovereign ruler. Hence, unlike Bacon, Hobbes does not endorse universal or philosophical language projects.

John Locke (1632–1704)

In Book III of *An Essay Concerning Human Understanding* (1689:1975) Locke develops Hobbes's line of linguistic thought into a full-fledged philosophy of language that arguably is not just "the first modern treatise devoted specifically to the philosophy of language," but the first treatise devoted specifically to the philosophy of language (Kretzmann, 1968: 175–6; Losonsky 2001: 5–21). He maintains that questions about what there is as well as questions about the extent of human knowledge cannot be answered properly without understanding the nature of natural language, particularly the meaning of natural language. Locke decided to write a book with the title *Of Words* because he found "that it is impossible to speak clearly and distinctly of our Knowledge . . . without considering, first, the Nature, Use, and Signification of Language" (1975: 401). Language became so important in Locke's mind, that he recommended that one of the three branches of science is semantics: "*sēmeintikē*, or *the Doctrine of Signs*, the most usual whereof being Words" (1975: 720).

Locke followed Hobbes in locating the meaning of language in the human understanding: "Words in their primary or immediate Signification, stand for nothing, but the Ideas in the mind of him that uses them" (1975: 405). The meaning of language, then, rests on ideas that are internal to the individual speaker's mind. To be sure, words can also refer to natural objects, for example water or gold, but this reference is only secondary and mediated by the ideas that words signify. For Locke, human beings succeed in referring to objects external to their minds only if they have ideas that refer to those objects, and it is an abuse of language to assume that words directly refer to external objects without the mediation of human ideas (1975: 497).

Locke assigned the study of language such a central role because he concluded that language plays a constitutive role in the formation of the ideas of the understanding. A problem for the meaning of language is that the ideas of the human understanding are unstable and human beings need greater order and stability than is found in human experience (1975: 464). Locke believed that human conscious experience begins with a rich but confusing array of individual simple ideas, for example ideas of particular shapes and shades of color, but human beings need complex ideas, for example, ideas of objects. According to Locke, without language the human understanding would not be able to regularly think about these ideas: " 'Tis the Name which is, as it were the Knot, that ties them fast together" (1975: 434).

Complex ideas that require language, according to Locke, are ideas of sorts, species, or kinds that have necessary and sufficient conditions. Neither nature nor human experience has such sharp boundaries, and it is only by abstracting certain ideas and annexing a name to them that human beings are able to form sortal ideas with necessary and sufficient conditions. Locke believes that moral ideas also require language (1975: 432 and 479).

For Locke all ideas are private—they are features of our individual caves, to use Bacon's terminology—and so human communication is a source of doubt and uncertainty, especially when it comes to complex ideas (1975: 476–7). Human beings cannot simply assume that speaker and hearer have the same ideas. Many words for complex ideas that individuals have cannot be translated (1975: 432–2). At best, speakers and hearers can aim to signify similar ideas with their words, and this requires effort in the form of the rectification of ideas. Rectification is the process for determining that two human beings have ideas that conform to each other by trying to fix ideas to a standard in nature. While rectification improves communication, it is always fallible and never yields absolute certainty (Losonsky 2007: 292–6).

Needless to say, Locke did not think that a universal or philosophical language was possible. While natural languages can be improved by being careful about the signification of words, the idea of a perfect language that avoids the imperfections of language is "ridiculous" (1975: 509). Without a system of ideas that is stable across time and shared by all human beings, a universal language, as long as it is assumed that the meaning of natural language is determined by the ideas of the human understanding, is impossible. Not even formal logic can provide an underlying unity. Formal logic is "useless" and just a "Web of perplexed Words" that undermines natural language, which is "the Instrument and Means of Discourse, Conversation, Instruction, and Society" (1975: 494–5).

Locke's philosophy of language is a culmination of a development that begins with a line of thought that emphasizes the diversity of human ideas, the dependence of the meaning of language on the ideas of the human understanding, the artifactual nature of natural language, the resulting diversity of human languages, and the role language plays in structuring the human understanding. For Locke, this means that philosophers must pay attention to human language when addressing metaphysical and epistemological issues, and more generally, human beings, including scientists, must curtail their aspirations to knowledge and temper their confidence even where knowledge is possible.

Linguistic Unity

René Descartes (1596–1650)

The responses just sketched make clear that in the modern period it was no longer possible, without additional justification, to assume a reliable correspondence between the human mind and reality. Nevertheless, the Aristotelian assumption that linguistic unity rests on an underlying psychological unity continued to thrive in seventeenth-century pursuits of a universal language as well as the allied pursuits of a universal grammar or logic.

An influential moment in seventeenth-century linguistic universalism is Descartes' letter in 1629 (1991, III: 10–13) responding to Marin Mersenne (1588–1648), who had raised the question of prospects for universal language projects. Descartes replies that universal language projects are possible only if all human ideas have a combinatorial structure like the set of natural numbers. Just as an infinite set of numbers can be known on the basis of a finite set of numbers and operations, a universal language would require that all human ideas are "compounded" out of a finite set of simple ideas and a finite set of operations. Given this structure of ideas, the universal language would then be mapped onto this order of ideas in the way that the language of arithmetic corresponds to the set of natural numbers. While Descartes dismisses the suggestion that such a

universal language will be widely adopted and used, he believes that such a universal language is possible because human ideas indeed have a combinatorial structure. The discovery of such a structure would require a "true philosophy," but Descartes believed that this is attainable.

Unlike Bacon, Hobbes, and Locke, Descartes does not worry about the instability of human thinking. What is unstable, Descartes maintains, is human imagination and sensation, and those who argue for psychological diversity confuse ideas with images and sensations. Ideas, Descartes maintains, are pure mental conceptions that even a disembodied being without images and sensations can have, and these mental conceptions have a combinatorial structure on their own independently of language. Moreover, Descartes maintains with Aristotle that different human beings can have the same mental conceptions: "A Frenchman and a German can reason about the same things" (1991, II: 126).

Finally, even though natural languages are ambiguous and irregular, they nevertheless succeed in representing the mind's combinatorial structure. In the *Discourse on the Method of Rightly Conducting One's Reason and Seeking the Truth in the Sciences* (1637) Descartes writes that with language human beings express their thoughts insofar as they are able to "produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in [their] presence" (1991, I: 140). Descartes believed that only rational beings can have a language with a combinatorial structure, and thus possession of language distinguishes beings with reason from beings without reason, including nonhuman animals (1991, I: 140–1).

Antoine Arnauld (1612–1694) and Claude Lancelot (1615–1695)

If reason is common to all human beings and has a determined combinatorial structure that is captured by the grammatical structure of natural languages, then there is a grammatical structure that is common to all human languages—a universal grammar. This is the hypothesis that was developed by Arnauld, a sympathetic reader and critic of Descartes, together with the grammarian Lancelot in *General and Rational Grammar* (1660), commonly referred to as the *Port-Royal Grammar*. In the Preface Lancelot writes that his work on Greek, Latin, Italian, and Spanish grammars led him to questions about the unity and diversity of the grammars of natural languages, and Arnauld provided him with a strategy for handling these questions.

A key idea of the *Grammar* is that language's primary purpose is to express thoughts, which have a structure that can be examined and understood without words or any other signs, and in order to serve this purpose, language has a grammatical structure in common with the universal structure of thought (1975: 66). The main aim of Arnauld's and Lancelot's *Grammar* is to explain grammatical structure in terms of the structure of thought. Since they assumed that the structure of thought is rational, the universal grammar was also a rational grammar.

The pursuit of a universal grammar alone does not distinguish the *Grammar* from Scholastic predecessors. The basic strategy of Scholastic universal grammars was to describe grammatical categories such as *noun* or *gender* in terms of properties of real objects (Law 2003: 173–77). This assumes that human conceptions of these objects are reliable, but early modern philosophers were under severe pressure to worry about reliability. Accordingly, what is new about Arnauld's and Lancelot's approach is that they tie linguistic categories explicitly to mental categories rather than ontological cat-

egories. Arnauld and Lancelot develop their mental categories on the basis of a fundamental distinction between the "object of our thought" and the "form or manner of our thought," both of which "occur in our minds" (1975: 67). The *object of thought* is the conceptual content of thought, and content is literally something that is grasped or intuited by the mind, even in cases where we are thinking about something that does not exist. Under the heading *manner of thought* Arnauld and Lancelot include propositional functions such as conjunction and disjunction, as well as propositional attitudes such as a belief, desire, question, or command.

The various kinds of objects of thought and manners of thought Arnauld and Lancelot identify are mental categories, not linguistic categories. For example, when they discuss *propositions*, *subjects*, and *predicates*, Arnauld and Lancelot assume that these have a status that is independent of language and that they can be known without language. However, natural languages, as instruments for expressing our thoughts, have corresponding grammatical categories. Nouns, articles, pronouns, prepositions, and adverbs, for example, are grammatical categories for terms that signify the objects of thought while verbs, conjunctions, and interjections are grammatical categories for words that signify the manners of thought. Arnauld and Lancelot then proceed in a familiar Aristotelian fashion, for instance, explaining grammatical nouns and adjectives in terms of the distinction between substance and accident. However, the substance/accident distinction in the *Grammar* is not intended to be a metaphysical distinction that carves out what there is, but a distinction between different kinds of ideas or conceptual contents (1975: 73). Human beings have *concepts* of substances and attributes, and Arnauld and Lancelot assume with Descartes that human beings can know what concepts they have and how they are related to other concepts without assuming that we know whether or not these concepts correspond to or represent anything real.

Arnauld's and Lancelot's *Grammar* is a self-contained treatise devoted to the nature of the meaning of natural languages, and insofar as the philosophy of language is concerned with semantic questions, this is an important text in the history of the philosophy of language. However, there is a sense in which Locke's Book III of the *Essay* remains "the first modern treatise devoted specifically to the philosophy of language." Arnauld's and Lancelot's discussion of grammar is not intended to make a broader contribution to philosophy in the way Locke's discussion of language contributes to his epistemology, metaphysics, and philosophy of mind. Their discussion of language assumes a broadly Cartesian conception of the mind, but they do not even make the minimal claim that the *Grammar* confirms or corroborates their conception of the nature of the human mind.

John Wilkins (1614–1672) and Jakob Böhme (1575–1624)

While Arnauld and Lancelot find the unity in the diversity of languages in a universal and determinate mental structure, other seventeenth-century universalists aimed to tie language directly to objects. There were two main approaches for preserving the Aristotelian view that languages are tied to objects. One approach took its cue from the natural sciences while the other approach assumed that there was a *language of nature* that expressed the essences of objects.

Language of nature doctrines were tied to the religious belief that the language of nature is the language God used in creation and gave to Adam. They were also combined with a disdain for human volition. But there were other features of the language

of nature doctrines that were equally important, including that language is not conventional, that it is acquired by human beings simply in virtue of being human without learning or instruction, and that language naturally and accurately represents reality.

A major source for seventeenth-century language of nature doctrines was the widely translated work of Böhme, who believed that certain sounds correspond to the essential properties of objects, and that these sounds comprise a language that can be understood simply by hearing the sounds. In England this view was defended in 1654 by John Webster (1610–1682) in a widely read pamphlet calling for the reform of university education (Debus 1970). Webster laments that the language of nature, which is "not acquired by us, or taught us by others," but "infused" and "innate," is neglected in the curriculum (26).

A pamphlet defending the university curriculum and attacking the very idea of a language of nature was published by Wilkins and the mathematician and astronomer Seth Ward (1617–1689). They rejected the assumption that what human beings acquire by their "own industry" is essentially deficient (Debus, 3–4). The natural sciences are a product of human industry, as are complex ideas and human languages. Words are always tied to objects by voluntary human decisions.

Wilkins was the author of the best-known attempt to develop a universal language, his *Essay Towards a Real Character and a Philosophical Language* (1668). Wilkins, a cofounder of the *Royal Society*, dedicated to the "Improvement of Natural Knowledge," wanted to develop a language that would serve the advancement of science. Wilkins' *Essay* is a prolegomena or preparatory study for a universal and philosophical language. The key idea is that this language consists of a finite and unambiguous set of simple terms tied to simple ideas. These terms can be combined using a finite set of rules to form more complex terms and sentences. The vocabulary and rules of this language correspond to ideas and structures of the human mind that are common to all human beings. Wilkins assumes, following Aristotle, that human beings have "the same Principle of Reason" and "the same *Internal Notion* or *Apprehension of things*" (1968: 20).

This universal language is also a philosophical language because the ideas that this language expresses correspond to reality. Words will "signify things" and the relationships between these words, as expressed in definitions, will be "suitable to the nature of things" (1968: 21). The reason the terms of this language refer to objects is that the terms correspond to internal notions that correspond to reality. Although this appears to rehearse Aristotle's view that the affections of the soul correspond to real objects, Wilkins was a product of the scientific revolution and he acknowledged the epistemic concerns raised by it. Wilkins aimed to tie his language to reality by constructing it on the basis of a "*Theory* . . . exactly *suted to the nature of things*" (1968: 21). For Wilkins, a theory that best fits the nature of things is a scientific theory, that is, a theory constructed according to scientific method, based on empirical evidence, and verified by the scientific community. For Wilkins, science is the source for the knowledge of things as well as their structure, including the predicamental structure he uses, that is ordering everything in terms of genus, difference, and species.

Wilkins was well aware that his classification was defective and only provisional, and that it would be improved over time and with scientific cooperation (1968: Epistle and 21). This suggests that Wilkins's method should be seen as anticipating strategies of contemporary naturalism which holds that the natural sciences are appropriate sources for information about what there is, and that philosophy, including the philosophy of language, should rely on this information.

Gottfried Wilhelm von Leibniz (1646–1718)

Leibniz offers an alternative strategy for tying words to things, namely relying on the correspondence between relationships. A language will have "a kind of complex mutual relation or order" and this linguistic structure can correspond to the structure of things (Leibniz 1970: 184; see Losonsky 2001: 64–5). Two distinct structures can have a one-to-one correspondence in the way a circle and ellipse can correspond to each other: for every point on a circle there is a point on the ellipse and the points are ordered in the same way. Thus human beings do not need to rely on a conventional connection between language and reality. If language represents reality, it will be in virtue of corresponding structures, and according to Leibniz whether these correspondences obtain is a nonarbitrary and nonconventional matter of fact.

By relying on corresponding structures, Leibniz offered a new way of identifying the universal structure of all languages. There is a universal linguistic structure if the diverse structures of natural languages correspond to each other. An important consequence of this approach is that human beings need not be aware of structures and their correspondences. Responding to Locke's view that human thoughts are private and fleeting, Leibniz distinguished between "actual thoughts" or "noticeable thoughts," on the one hand, and the "very form or possibility of those thoughts" (1996, 300–1). Actual thoughts are the conscious thoughts a person has at a time, and these passing thoughts are typically momentary and private. But fleeting conscious psychological states can also have a stable form that need not be noticed by the person in those states. The search for universality, then, is not a search for some conscious state that all human beings have but for the *form* of those states.

One form that concerns Leibniz is semantic content. Actual thoughts have content, or what Arnauld and Lancelot called "objects of thought." Leibniz identifies semantic content in terms of logical relations, specifically logical entailment. One important feature of a proposition is its logical relations to other propositions: what it entails, what it does not entail, what it contradicts, and so forth. These logical relations are not constructed—Leibniz describes them as the "natural order of ideas"—and they are shared if there is a correspondence between the ideas of different people. Moreover, insofar as sentences of natural languages have semantic content, they also have a universal structure, namely their logical form [*forme logique*] (1996, 480). Thus the search for a universal structure for all natural languages is the search for a universal logical form of all natural languages.

Leibniz identifies two tools for uncovering the logical form of natural language. The first is a logical calculus, which he began developing in his *Dissertation on the Art of Combinations* (1666). Adopting Hobbes's suggestion that all reasoning is computation, Leibniz actually develops a calculus that represents mathematically the formal structure of valid inferences. With this calculus it is possible to exhibit the entailment relations between sentences and thus exhibit their logical form. The second tool is the substitution of semantically equivalent terms and phrases. Leibniz believes that an orderly sequence of substitutions of equivalent terms and phrases can transform an ordinary sentence of a natural language into a more precise sentence that uses the logical calculus to explicitly express the logical structure of the original sentence. This transformation or translation counts as an analysis of the original sentence. Leibniz also offers a tool for identifying equivalent terms: "Those are the same of which one can be substituted for the other with truth preserved," namely the *Principle of Substitutivity Salva Veritate* (1966, 34, 53 and 122).

However, for Leibniz there is more than just a correspondence between language and mind. Leibniz departs from the orthodox Aristotelian and Cartesian conception of human language as a mirror of the mind and agrees with Hobbes and Locke that natural languages and other symbol systems have a constitutive role to play in human reasoning. Leibniz maintains that human beings cannot reason without symbols (1996, 77 and 212; 1966b, 450). Just as a person can calculate quickly and successfully by writing numerals and function symbols on a sheet of paper without being conscious of the meaning of the numerals, human beings think when speaking even though they do not have an additional idea or image. This is not only true of public speech, but when thinking privately without speaking, the train of thoughts can consist of ideas of words (1966b, 451). Although language plays a constitutive role in human thinking, this does not undermine the thesis that there are universal components to human thought because language can express a universal order, namely a logical order.

Conclusion

Seventeenth-century philosophers of language developed several basic themes that endured as the philosophy of language developed in subsequent centuries. Arguably, there is a tradition beginning with Locke that emphasizes linguistic and psychological diversity and gives increasing importance to the role human language has in structuring the human mind (Aarsleff 1982). Linguistic diversity and relativity characterize the eighteenth-century philosophies of language of Étienne Bonnot Condillac (1715–1780) and Johann Gottfried Herder (1744–1803). However, in the eighteenth-century emphasis shifted from individual psychological differences to broader social psychological and anthropological differences. For example, Condillac maintains that nations have different characters that shape language even more than individual characters, and this yields the special character or "genius of a language" (2001, 185–195). This trend continues in the nineteenth century, especially in the work on non-Indo-European languages by Wilhelm von Humboldt (1767–1835).

The pursuit of a universal grammar also continues in the eighteenth century. Horne Tooke (1736–1812) aimed to combine Locke's philosophy of mind and language with a universal grammar. James Harris's (1709–1780) *Hermes or a Philosophical Inquiry Concerning Language and Universal Grammar* (1751) aims to combine a universal grammar with anthropological concerns. Not only do individuals have ideas, Harris writes, but also nations have ideas—"the genius of their nation"—and some nations have better ideas and better-organized ideas than others (1751, 407–8).

While Leibniz's logical calculus based on mathematical functions influenced eighteenth-century mathematicians Leonhard Euler (1707–1783), Gottfried Ploucquet (1716–1790), and Johann Heinrich Lambert (1728–1777), the project of exhibiting the logical form of natural languages was dormant in the eighteenth century. For instance, Immanuel Kant's (1724–1804) interest in the logical form of human cognition did not turn into a concern for the logical form of natural language. This concern reemerges in the nineteenth century in the writings of Bernard Bolzano (1781–1848), John Stuart Mill (1806–1873), and Gottlob Frege (1848–1925). All three isolated the proposition as a unit of logical and linguistic analysis and distilled the features of propositions that are relevant to logical validity. However, Bolzano and Frege, acknowledging Leibniz's influence, sharply separate logic and psychology, while Mill follows Locke in identifying logical relations with psychological operations.

Despite their differences, Bolzano, Mill, and Frege all agree with Locke that the problems of philosophy cannot be solved without a careful study of language and its meaning. Insofar as Bolzano, Mill, and Frege are major sources of twentieth-century and contemporary philosophy of language, particularly the philosophy of language in the analytic tradition, philosophy of language today is rooted in the turn to language by seventeenth-century philosophers.

References

Aarsleff, H. (1982) *From Locke to Saussure: Essays on the Study of Language and Intellectual History*. University of Minnesota Press.

Arnauld, A. and Lancelot, C. (1975) *General and Rational Grammar: The Port-Royal Grammar*. Translated by J. Rieux and B. Rollin. Mouton.

Bacon, F. (2000) *The New Organon*. Translated by L. Jardine and M. Silverthorne. Cambridge University Press.

Bacon, F. (2001) *The Advancement of Learning*. Ed. S. J. Gould. Random House.

Condillac, E. B. de. (2001) *Essay on the Origin of Human Knowledge*. Cambridge University Press.

Debus, A. G. (1970) *Science and Education in the Seventeenth Century: The Webster-Ward Debate*. American Elsevier.

Descartes, R. (1991) *The Philosophical Writings of Descartes*. 3 Vols. Translated by J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny. Cambridge University Press.

Harris, J. (1751) *Hermes: or, A Philosophical Inquiry Concerning Language and Universal Grammar*. Woodfall.

Haugeland, J. (1989) *Artificial Intelligence: The Very Idea*. MIT Press.

Hobbes, T. (1981) *Computation sive Logica: Logic*. Translated by I. C. Hungerland, A. Martinich, and G. R. Vick. Abaris Books.

Hobbes, T. (1996) *Leviathan*. Cambridge University Press.

Kretzmann, N. (1968) "The Main Thesis of Locke's Semantic Theory." *Philosophical Review* 77: 175–96.

Law, Vivien. (2003) *The History of Linguistics in Europe: From Plato to 1600*. Cambridge University Press.

Leibniz, G. W. v. (1966a) *Logical Papers*. Translated and edited by G. H. R. Parkinson. Clarendon Press.

Leibniz, G. W. v. (1966b) *Deutsche Schriften*. 2 vols. Ed. G. E. Guhrauer. Olms.

Leibniz, G. W. v. (1970) *Philosophical Papers and Letters*. Ed. L. Loemker. Reidel.

Leibniz, G. W. v. (1996) *New Essays on Human Understanding*. 2nd ed. Translated by P. Remnant and J. Bennett. Cambridge University Press.

Lewis, R. (2007) *Language, Mind and Nature: Artificial Languages in England from Bacon to Locke*. Cambridge University Press.

Locke, J. (1975) *An Essay Concerning Human Understanding*. Oxford University Press.

Losonsky, M. (2001) *Linguistic Turns in Modern Philosophy*. Cambridge University Press.

Losonsky, M. (2007) "Language, Meaning and Mind in Locke's *Essay*." In *The Cambridge Companion to Locke's "Essay Concerning Human Understanding."* Ed. L. Newman. Cambridge University Press, 286–312.

Wilkins, J. (1968) *An Essay Towards a Real Character and Philosophical Language*. The Scolar Press.

7.4
FREGE, RUSSELL, AND WITTGENSTEIN

Michael Potter

Frege, Russell, and Wittgenstein were founders (although not by any means the sole founders) of the analytic tradition in philosophy; but they did not found the philosophy of language, which has roots stretching back much further. Their principal contribution, indeed, could be regarded as being in the opposite direction: it is not so much that they applied philosophical methods to the study of language as that they applied linguistic methods to the study of certain problems in philosophy. In the course of this work they did develop ideas which shed light on language and how it functions. However, even this must be heavily qualified, since their main contributions were much more to the philosophy of logic (the study of the inferential role of sentences) than to the philosophy of language (the study of how language means what it does). In the summary that follows we shall focus on the morals that can be drawn from their contribution for modern work in the philosophy of language.

Frege

It is worth noting straightaway that mathematics played an important role in shaping the philosophical ambitions of all three of the philosophers we shall be considering. Frege was a mathematician for all of his professional life; Russell began as one and much of what he wrote had the philosophy of mathematics as its focus; and Wittgenstein, although he did not train as a mathematician, announced when he arrived at Cambridge in 1911 that it was the philosophy of mathematics that he wanted to study with Russell. The desire to secure the philosophical foundations of mathematics which all three of these philosophers shared plays a large part in explaining the attitude to language which—to begin with, at least—they chose to take. Frege began with the ambition of formulating a precise language—a *Begriffsschrift*, he called it—for the expression of thoughts. Where the thoughts to be captured are mathematical, Frege's ambition was an entirely plausible one, which he went some way towards realizing. It was natural, once the method had been shown to be successful in mathematics, that he and others should consider applying it more widely. However, the difficulties of formalization are much more severe where nonmathematical language is concerned, and the relationship between natural languages and their formalized correlates remains problematic to this day.

Perhaps the most remarkable of Frege's contributions to philosophy was what is nowadays called the linguistic turn (a term popularized by Rorty but coined by Gustav

Bergmann). This was the movement that placed a concern with language at the center, not only of the philosophy of mathematics but of all of philosophy. What made it central was the realization that we approach the world via thought, but have no access to thought except via language. Hence, it was claimed, our inquiry into the nature of the world is best conducted via an inquiry into the nature of language.

One of the most important contributions made by Frege was to place language at the center of philosophical, and in particular metaphysical, inquiry by recognizing its importance as a route to the structure of our thinking about the world. This approach was encapsulated in what has become known as the context principle, namely the injunction that we should not enquire after the meaning of a word in isolation but only in the context of a sentence. Frege formulated this principle in order to apply it to the case of number-words, and hence to rule out the question as to what the numbers are, posed independently of the question how number-words are used in sentences. Applied more generally, the context principle led during the twentieth century to an approach which has become known as internalist. This approach seeks to discuss metaphysical questions about the structure of the world by means of a discussion of the structure of the language in which we represent the world.

If it were only a route to internalism, the context principle would not amount to an important contribution to the philosophy of language. However, it has also been influential within that area, as it has led to an extended debate about the overall shape that theories of meaning should have. Applied in its strongest form, the context principle would require that to understand a single word I need to understand the whole of language, and hence that each time I learn a new word my understanding of every other word must change slightly. Donald Davidson (1967, p. 308), for instance, has advanced the context principle in this strong form. Many twentieth-century advocates of the context principle, of whom Michael Dummett is the most notable, have preferred to limit its application so that it does not have this holistic consequence.

Another Fregean contribution which has influenced the philosophy of language is his notion of thought and, derivatively, of sense. Thoughts are not, according to Frege, mental entities in the mind of an individual thinker: they are intersubjectively available and hence cannot be wholly any one person's property. It is worth stressing here that the publicity of thoughts is not itself a distinctively Fregean contribution. What *is* distinctive is that Frege combined this with a second element, namely that what he meant by a thought was only that part of the content of the sentence that is relevant to inference. The residue, which he called "tone," is no doubt important for the philosophy of language but is not the concern of logic. By means of this restriction Frege hoped to explain the independence of logic from the mind. His target, that is to say, was in the first instance the view not that language is psychological but that logic is. What he railed against was the view, popular in the late nineteenth century, that logic is a codification of how we think. Frege held that this is utterly wrong-headed: logic, he claimed, codifies the laws of truth, not of thought; the laws of how we *ought* to think if we wish to aim at the truth, not of how we do actually think.

Much of the later interest in Frege's notion of sense has focused on its application at the subsentential level, to proper names in particular. In "On sense and reference" Frege argued that in order to explain the informational content of identity sentences such as "Hesperus=Phosphorus" we cannot appeal merely to what the two sides of the equation refer to, since this is the same in both cases (namely the planet Venus). Nor can we explain the content solely in terms of the difference between the *words* "Hesperus" and

"Phosphorus," since that would, according to Frege, make the content linguistic rather than, as it should be, astronomical. Nor can we base our account on the difference in the ideas which I associate in my mind with the two words, since these, being mental, are intrinsically private, whereas the informational content of the equation is something public—something which you can learn from me when I tell it to you. For these reasons Frege felt driven to the conclusion that the contributions of the words "Hesperus" and "Phosphorus" to the meanings of sentences in which they occur—what he called the *senses* of the words—are different, but are inter-subjectively available items which are not to be identified either with the words themselves or with anything mental.

In "On sense and reference" (1980a) Frege motivates the notion of sense for proper names primarily on the basis of the need to explain intersubjective communication. The notion that there could be such a thing has seemed puzzling to many philosophers since then, from Russell to Kripke and beyond. When asked some years later to justify his view, Frege invoked (1980b, p. 80) the example of two explorers, one of whom comes across a mountain called Aphla, the other a mountain called Ateb. Only when they plot their discoveries on a map do they come to realize that Aphla and Ateb are the same mountain, but seen from different directions. In the form in which Frege told the story, it seems to make again the point about publicity: the notion of sense is required if we are to explain how communication between individuals is possible. It is interesting to note, however, that the problem arises even if we revise the story so that only one explorer is involved. On this retelling, the moral of the story seems to concern the nature of the world as much as the structure of thought. It is because the world has unexpected and unknown aspects that we need a notion of sense capable of making identity statements nontrivial.

I have said that Frege's principal concern was with the philosophy of logic, not of language. In a few places, though, Frege did go beyond these narrow bounds. This is particularly notable in "On sense and reference," where the discussion of indirect speech goes some way beyond anything that would have been necessary if his interest had been strictly limited to the foundations of mathematics.

One ongoing theme in Frege's writings is the complicated relationship between language and thought. In his late writings he seems largely to have abandoned the idea that language was a good guide to the structure of thought, but his reasons are not wholly clear. The most obvious one might be his paradox of the concept *horse*, but he did not himself take this paradox as seriously as he might have done: It was left to Wittgenstein to do that. More important for Frege seems to have been the failure of his logicist project for grounding arithmetic. He seems to have thought that this was somehow due to a mismatch between language and thought, but he was unable to articulate the nature of the mismatch with any clarity.

Russell

What is often taken to be Russell's most famous contribution to the philosophy of language is his theory of descriptions (Russell, 1905), according to which sentences containing definite descriptions should be analyzed so that the description disappears and its role is taken over by quantifications. However, Russell did not originally intend this theory as a contribution to the philosophy of language at all, and for two reasons. The first was that, like Frege, he was at this stage more interested in the logical rather than the grammatical form of propositions. The difference in form between "John is bald"

and "The present King of France is bald" is one which emerges when we consider how these sentences behave when we negate them. The latter sentence has two negations with different truth conditions ("The present King of France is not bald" and "It is not the case that the present King of France is bald"), whereas the former only has one negation. Russell conceived of this difference as a logical rather than a grammatical one. For him, therefore, the theory of descriptions was a demonstration that the logical and grammatical forms of a sentence may be different.

The second reason that Russell did not see himself as contributing to the philosophy of language is one that emerges when we recall the reason why he was interested in the King of France's hair in the first place. For him this was a proxy for the much more important case of the class of all classes which do not belong to themselves. In other words, if he was concerned with language at all, it was only the language of mathematics, not that of ordinary life. So his aim was not to give an account of all the various uses of the word "the" in ordinary language. The fact that his theory does not cope well with "The whale is a mammal" did not seem to him to be an objection to it. How far Russell was from conceiving of "On denoting" as a contribution to the philosophy of language may be gauged from his response to Strawson's "On Referring." Strawson had criticized the theory of descriptions because there are various features of our use of definite descriptions which it fails to analyze correctly. Russell's reply was that he was "totally unable to see any validity whatever in any of Mr Strawson's arguments" (Russell, 1957, p, 385).

During the most philosophically influential part of his working life—the decade or so immediately preceding the First World War—Russell did not begin by being interested in language in its own right at all. Nonetheless, "On denoting" did have the effect of forcing him to consider the role of language, even within mathematics, more carefully than hitherto, because it showed him that the surface grammar of a sentence may mislead us as to the true form of the thought the sentence expresses. He came to realize, as Frege had by a different route, that if language is a medium for the expression of thought, it is not a wholly transparent one.

Russell's work during his most productive phase contributed to a developing research project which might be thought of as an analysis of language but was heavily influenced by epistemological concerns deriving from Descartes. The underlying language, that is to say, was to consist of elements corresponding to whatever is most immediate in experience. The project then consisted in the construction, from these elements, of linguistic items corresponding to the objects of the external world as we ordinarily conceive of it. Even when Russell gave more explicit attention to language in his later work (such as the *Inquiry*), the approach he took was shaped by the epistemological concerns of his project, thus limiting the interest of this work to anyone who does not share these concerns.

The *Tractatus*

The *Tractatus* is famous for its advocacy of a so-called picture theory of language, but this is at best a misnomer, since what constitute pictures of the world are, according to the theory, not the sentences of ordinary language but the propositions which those sentences express. For this reason "picture theory of thought" would be a less misleading title for what Wittgenstein intended.

According to the picture theory what enables a picture to represent a situation is that the structure of the picture is identical to the structure of the situation. If the aim of

semantics is to explain how propositions come to have meanings, then according to the picture theory there is only one task which semantics needs to perform, namely that of connecting the components of the proposition with the components of the world. The further task of explaining how the way in which the components are assembled succeeds in being expressive of a certain situation drops away, since it is just the same way that the components of the world are assembled if the proposition is true. For Wittgenstein, therefore, the reason why thought succeeds in latching onto the world is that its structure is identical to that of the world.

I have stressed that it is not the sentence but the proposition (or, in Frege's language, the thought) which according to the picture theory shares its structure with the world. In order for sentences also to share this structure it would be necessary for us to construct a language which mimicked it. In the *Tractatus* Wittgenstein went some way towards constructing such a language, although at times he seems to have been more concerned with explaining how Frege's and Russell's formal languages fell short of this ideal than in the details of how his own construction was supposed to work.

The idea that the world is one that is represented in thought is at the heart of Wittgenstein's atomism, since he moves (1992, 2.021–2.022) from an argument that determinate thoughts are atomistic to the conclusion that the world is correspondingly atomistic. Moreover, in a further simplifying move he claimed that these atoms are simple not only in the sense that they have no parts but also in that they do not have distinct aspects: An atom which is presented to me in one way cannot later be presented to me in another. By this means Wittgenstein ruled out the possibility which Frege had used the example of Aphla and Ateb to highlight. From this Wittgenstein derived his essentially solipsistic conclusion that there cannot be different perspectives on the world.

One theme which lies just below the surface of much of Frege's philosophical writing is the relationship between thought and the world. Frege took it as given that the structure of thought mirrors the structure of the world, so that the distinction between saturated and unsaturated components of the thought corresponds precisely to a distinction between saturated and unsaturated components of the world. However, Frege's conception of objects prevented him from holding that this neat isomorphism holds within these grammatical categories. Wittgenstein, by rejecting Frege's notion of objects, was able to extend Frege's conception of thought mirroring the world much further.

The picture theory can thus be seen as representing the high point of Frege's linguistic turn as a method in metaphysics, since it holds out the hope that if we constructed an ideal language, and thereby uncovered the structure of thoughts, we could then read off the structure of the world directly from them. However, the logically perfect language is offered in the *Tractatus* more as an ideal than a practical reality, and by the time he finished the book he had begun to doubt whether the ideal was realizable at all.

By sidelining ordinary language in this manner Wittgenstein left rather obscure what philosophical purpose is served by studying it. He said that "all propositions of our colloquial language are actually, just as they are, logically completely in order"(1922, 5.5563), but it is an interesting question quite what he meant by this, and it is a tension that remains unresolved in the *Philosophical Investigations* (see Wittgenstein 1953, §98).

Nonetheless, Wittgenstein still advanced in the *Tractatus* the ambition of using the analysis of language to solve philosophical problems. He proposed a model according to which most of the things we ordinarily say could in principle be expressed in the ideal language. Some things we say resist translation into this ideal language and are thus revealed to be nonsense masquerading as sense. In particular, the *Tractatus* problematizes much of

the philosophy of language, since its explanations of how language comes to have meaning inevitably make use of the very features of language which it seeks to explain.

One instance of this is the notion of a grammatical category. In later life Wittgenstein was fond of quoting, as a guide to good philosophizing, a maxim of Hertz to the effect that our minds will "cease to ask illegitimate questions" when we remove unnecessary variables from our specification of a problem. In the *Tractatus* we can see Wittgenstein applying this maxim in order to reject the conception of grammatical categories as akin to pigeonholes into which words can be sorted (a conception which he accused Russell of holding). According to Wittgenstein the "grammatical category" to which a word belongs is a label for certain structural features of its meaning, not something which can be separated from the word and discussed in isolation from it.

Later Wittgenstein

One way of summarizing the difference between early Wittgenstein and late is that in his later work he abandoned hope of a certain sort of tidying up of language. So what in the *Tractatus* is described as "*the* language which alone I understand" (5.62) is replaced by a motley of overlapping language games. Whatever the role of ideal language really is in the *Tractatus*, it is clear that assertoric sentences are there taken as primary, whereas in the *Philosophical Investigations* what is emphasized from the start is how many and various are the roles that language has.

The later Wittgenstein also gave up on Frege's notion of sense, if by that is meant the idea of something quite determinate which a sentence, on a particular occasion of utterance, can be said to mean. In the *Investigations* Wittgenstein offers numerous examples intended to loosen our grip on this kind of notion of meaning. Modern philosophers of language influenced by Wittgenstein sometimes suggest that meaning is so sensitive to context as to make the project of supplying a theory to account for it almost hopeless.

Another issue which the later Wittgenstein addressed was the relationship between the private and the public. I mentioned earlier that Frege characterized the mental as private and emphasized, by contrast, the intersubjective availability of the content of our utterances, from which he deduced that these contents cannot be mental. What Wittgenstein did was to question what role the private realm of the mental can *ever* play in explaining linguistic meaning. Understanding the sense in which a private language is impossible is important, he thought, since it does a great deal to delimit our explanation of how public language is possible.

Indeed, one of Wittgenstein's main targets in his later work was the whole conception on which Russell's Cartesian project was based. He questioned, that is to say, the assumption that our knowledge of our own ideas is somehow more certain than anything else, and hence that the correct method of epistemological enquiry is to work from the inside outwards. This critique, if correct, is of methodological importance for the philosophy of language much more widely, since the Cartesian assumption is often used as an unargued premise to motivate the direction in which explanations proceed.

In the *Investigations* Wittgenstein also offered a critique of an "Augustinian" account of language, by which he meant an account which sees it as possible to build up a theory of linguistic meaning with names as the starting point. In a way, Wittgenstein's objection seems now to be a variant of the one he had against what he saw as Russell's tendency to think of grammatical categories as being like pigeonholes. Augustine, he says, "describes the learning of human language as if the child . . . already had a language,

only not this one" (1953, §32). The extent to which the child might be said to "have a language, only not this one" is of course something that remains controversial.

So far, though, all of this is largely negative. Was there anything positive which the late Wittgenstein had to offer? It is sometimes said that he advanced a "use theory of meaning," but on its own this tells us very little, since it is hard to see how a credible theory of meaning could fail to appeal to how we use words and hence be in some sense a "use theory." What Wittgenstein intended, however, seems to have been not so much a theory as a method. This may well go some way towards explaining the apparent negativity of much of his later philosophy (philosophy as therapy).

Another suggestion that has been made is that ordinary language philosophy is an application of Wittgenstein's methods. It was he, after all, who described his task as being "to bring words back from their metaphysical to their everyday use" (1953, §116). However, as a matter of history the influence of the later Wittgenstein on Oxford ordinary language philosophers such as Austin and Ryle seems to have been rather slight, and the commonality of their concerns somewhat superficial.

Morals

What morals do these authors offer for modern work in the philosophy of language? The issues we have been discussing are almost all related to the notion of content—whatever it is that is expressed by a piece of language. As we have noted here, their focus on logic led these authors to focus on narrow notions of content from which those aspects not relevant to inference have been excluded. Yet such narrow notions plainly do not tell the whole story. Moreover, even within the narrow scope of logic, we should take note of the difficulties our authors found when they tried to determine what this content is. It is curiously hard to find an appropriate notion of content that is stable enough to have a determinate structure capable of precise theoretical investigation.

It follows, too, that it is important not simply to assume that the link between language and thought, which the founding fathers highlighted, is a wholly transparent one. The starting point is the idea that a sentence of one language can have the same content as one in another. The task of the translator, one might say, is to preserve as much of the content as possible in the process of translation. But as soon as we suppose that there is a notion of content which can be preserved, we have a puzzle as to its structure, since languages differ widely in how they say things.

Another moral is the need for consideration of the role of outlying cases. Language is messy, and almost every generalization one cares to make has exceptions. What is not so clear is how seriously we should take the exceptions. Our attitude to them will presumably depend on what we take the purpose of our work to be. If we have metaphysical or epistemological pretensions, then perhaps a few counterexamples are not so troubling. If, on the other hand, our purpose is to explain language as it exists, any policy of ignoring counterexamples would need to be theoretically motivated. We need, that is to say, to decide what our purpose is first, and only then proceed to offer a theory.

Above all, though, we need to think through the implications of two points made by Wittgenstein but deriving ultimately from Frege: first, that it makes no sense to enquire about a grammatical type independent of its instances; second, that it remains one of the central challenges for the philosophy of language to explain how its account of meaning can be related to our own mental lives.

Bibliography

Davidson, D. (1967), 'Truth and meaning,' *Synthese*, 17: 304–23.
Frege, G. (1980a), "On sense and reference," in *Translations from the Philosophical Writings of Gottlob Frege*, edited. and translated by P. Geach and M. Black, 3rd edn, Oxford: Blackwell, pp. 55–78.
Frege, G. (1980b), *Philosophical and Mathematical Correspondence*, Oxford: Blackwell.
Russell, B. (1905), "On denoting," *Mind*, 14: 479–93.
—— (1957), "Mr Strawson on referring,", *Mind*, 66: 385–9.
Wittgenstein, L. (1922), *Tractatus Logico-Philosophicus*, London: Kegan Paul, Trench, Trubner.
—— (1953), *Philosophical Investigations*, Oxford: Blackwell.

Further reading

Heck, R., and R. May (2006), "Frege's contribution to philosophy of language," in E. Lepore and B. Smith (eds.), *The Oxford Handbook of Philosophy of Language*, Oxford University Press.
Potter, M. (2009), chapters 8 and 26 of *Wittgenstein's Notes on Logic*, Oxford University Press.

7.5
LOGICAL POSITIVISM AND QUINE

Sanford Shieh

In the middle third of the twentieth century logical positivism was a dominant strand of analytic philosophy. Positivism began with a group of philosophers, mathematicians, and scientists—associated with the Vienna Circle and the Berlin Society for Scientific Philosophy—interested in philosophical understanding of modern science. Under the influence of Ludwig Wittgenstein, the Vienna Circle, especially Moritz Schlick, Hans Hahn, and Rudolf Carnap, came briefly to espouse a form of Humean empiricism incorporating the techniques of the mathematical logic developed by Gottlob Frege and Bertrand Russell. On this view, call it *classical positivism*, language is not only a central subject of philosophy but also the principal instrument of philosophizing. Classical positivism, as popularized by Ayer (1936), became identified with positivism *tout court*. But the positivists themselves were aware of an array of difficulties with the classical doctrines, and soon moved away in a number of diverging directions. Carnap, in particular, reached a thoroughgoing deflationary and pragmatic anti-metaphysical view, advocating the replacement of traditional philosophy by the construction of formal languages. W. V. O. Quine, student and friend of Carnap, engaged in a more than decade long debate with Carnap over Carnap's mature views. Quine's arguments came to be widely accepted as a decisive refutation of positivism, thereby helping to end its influence in the analytic tradition. Recent work has increasingly shown how little their debate has to do with classical positivism, and how inconclusive Quine's arguments against Carnap are, thereby opening the way to a better appreciation of the depth and philosophical interest of their disagreements (see, *inter alia*, Awodey and Carus 2007, Creath 2003, Ebbs 1997, and Ricketts 2003).

Classical Tractarian Positivism

Positivism is famously associated with a verificationist theory of meaning, on which, in order for a sentence to be meaningful, it has to be associated with a method of verification that is ultimately based on sensory experience. The initial notoriety of the Vienna Circle stems from their use of this criterion of meaningfulness to argue that metaphysics is not false but cognitively meaningless nonsense. The verification principle comes from the Circle's commitment to empiricism, the view that sense experience is the only source of genuine knowledge about the world. But the distinctive logical form of empiricism advanced by the positivists arose from an apparent incompatibility between

empiricism and their attachment to modern science as the paradigm of genuine knowledge. A key problem is that mathematics is an integral part of science, but it is unclear how mathematical knowledge could originate in sense experience. Experience might always be different from the way it is, so any truth based on experience is contingent. In contrast, we have no clear conception of how mathematical truths might be false. The same point evidently holds of logical truths. Thus, mathematical and logical knowledge seem to be *a priori*, and so to require some faculty of rational intuition, paving the way to metaphysics. Indeed, when combined with verificationism, these features of logic and mathematics lead to the unwelcome result that they are no more cognitively meaningful than metaphysics. So, "anyone who wants to advocate a consistent empiricism must face up to" this problem (Hahn 1980: 40).

The Circle saw, in Wittgenstein's *Tractatus* (1922), the solution. On their reading of the *Tractatus*, sentences in general represent the world by expressing truth functions of elementary sentences, which, in turn, represent by picturing atomic facts. The truth-functional mechanism of linguistic representation guarantees the existence of *tautologies*, sentences that are true no matter what are the truth-values of elementary sentences composing them. Logic consists of tautologies, sentences that owe their truth, not to correct depiction of the world but to the nature of linguistic representation. Since logic is not about the world, our knowledge of logic does not rest on the sensory sources of genuine knowledge of worldly facts. Rather, it rests on knowledge of how we represent the world in language, of the rules or conventions (that we tacitly accept as) governing the use of language, in short, on knowledge of meaning. The apparent apriority and necessity of logic can now be taken to have a linguistic, rather than factual, origin.

Clearly, this Tractarian view of logic has to be extended for it to yield the desired consistent empiricism. It has to cohere with verificationism—this was accomplished by taking elementary sentences, the fundamental Tractarian unit of representation, to depict, and so be verifiable by, immediate sensory experience. And it has to be applied to mathematics—here the *hope* was to carry out a type of logicist reduction of mathematics, to tautologies rather than Frege's or Russell's formulations of logic. *Modulo* the execution of this project, the positivists had a consistent empiricism and a basis for rejecting metaphysics as neither verifiable nor tautologous.

Three sets of problems with this view were discussed in the Circle. First, exactly what notion of verification is relevant to meaning? Is it actual, or merely in principle? Is it conclusive, or merely probable confirmation? Is it tied to occurrences of sensations or observations of macroscopic physical objects? Eventually there seemed to be some convergence on verification as confirmation in principle. But it was never clear how verification contributes to the empirical meaning of dispositional and theoretical terms of science (see Hempel 1965).

Second, the main obstacle for Tractarian logicism is the presence of unbounded infinitary quantification in mathematics; it is, to say the least, unclear how such sentences can be finite truth-functional combinations of elementary sentences. A similar problem afflicts a Tractarian account of the statements of science: If to be meaningful they have to be finite truth-functions, then they can be no more than summaries of extant observations; this rules out laws, unbounded quantifications ranging over potential observations.

Finally, what is the status of the verification principle itself, and more generally, the products of positivist philosophizing? Are these sentences empirically contentful? Or are they tautologies? As Carnap puts it in the form of an objection, "[i]n consequence of

your view your own writings . . . would be without sense, for they are neither mathematical nor . . . verifiable by experience" (1935: 36). In the background here is the question whether classical positivism represents a genuine escape from metaphysics. The apparently meta-philosophical critique of the nonsensicality of first-order metaphysical sentences seems to rest on certain first-order claims about the nature of meaning. Should these purported claims turn out to be as nonsensical as traditional metaphysics, it's hard to avoid the conclusion that the positivist "ready to prove that metaphysics is impossible is a . . . metaphysician with a rival theory" (Bradley 1897: 1).

Carnapian Positivism

A Transitional View

Carnap attempted to work out the details of Tractarian positivism using David Hilbert's (1925) conception of infinitary mathematics as contentless sequences of signs whose only legitimate use is for deriving contentful finitary mathematical sentences. Initially, this conception was harnessed to solve the twin problems of unbounded quantification mentioned above. The role of the contentless was to be played by unbounded quantifications, and that of the contentful by tautologous truth-functions in the case of mathematics and nontautologous ones in the case of science. Technical difficulties facing the mathematical case eventually led Carnap to abandon the Tractarian framework altogether in favor of the Hilbertian one (see Awodey and Carus 2007: 176–9).

Language became conceived as a *calculus*, "a system of rules" (Carnap 1937b: 4) rather than a system for picturing extra-linguistic facts. The study of language thus conceived was at first called "meta-logic" and eventually "logical syntax." The model for this discipline is Hilbert's meta-mathematics. Hilbert conceives of the standards governing mathematical proof as relations among sentences considered as meaningless finite sequences of objects. Analogously, Carnap's *pure* syntax studies the rules constituting a calculus in terms of properties of and relations among "possible arrangements" of "elements . . . about . . . which nothing more is assumed than that they are distributed in various classes" (ibid.: 4–5). The rules of a language are divided into two main categories: Formation rules specify which sequences of elements are expressions of the language, transformation rules specify the inferential properties and relations among expressions.

The question now is, how does this conception of language fit with empiricism and with the supposed nonempirical legitimacy of logic and mathematics? One account proceeds in terms of an analytic-synthetic distinction (hereafter AS). *Analyticity* is Carnap's successor to the Tractarian positivist notion of tautology. Since language is seen as a set of rules, the semantic notion of sentences true in virtue of the mechanism of representation is replaced by a syntactic notion of sentences true in virtue of the rules of the language. Empirically meaningful sentences are *synthetic*. Synthetic sentences are divided into observational and theoretical: The former acquire their empirical content by direct sensory observation, the latter via observational consequences. Logic and mathematics are precisely the rules governing the derivation of observational consequences. So transformation rules are logico-mathematical axioms and rules of inference.

Meta-mathematics also yields an account of the status of philosophy. Hilbert's insight is that properties of and relations among finite sequences are precisely the subject matter of contentful finitary mathematics, so mathematical justification can be studied

mathematically. Analogously, Carnap's logical syntax—meta-logic—is the mathematical study of the rules of language. Once philosophy is "replaced by . . . logical syntax" (1937b: xiii), philosophy is mathematics, non-empirical but not nonsensical.

Logical syntax

Carnap's (1937b) (hereafter *Syntax*) contains a fundamentally different view. To begin with, transformation rules are not restricted to logic and mathematics: "[w]e may . . . also construct a language with *extra-logical rules of transformation* [including] amongst the primitive sentences the so-called laws of nature. . . . It is possible to go even further and include . . . observation sentences" (1937b: 180; hereafter all emphases are Carnap's unless otherwise specified). Moreover, Carnap does not even require the transformation rules to be given as divided into logical and physical rules. Purely syntactically specified transformation rules "determine . . . *the distinction between logical and descriptive symbols*, . . . and further, *between logical and extra-logical* (*physical*) *transformation rules*, from which . . . 'analytic' arises" (1937b: 168–9). So in *Syntax* AS is not merely language-relative but is grounded on primitive syntactic (inferential) characteristics of each language, rather than extra-linguistic metaphysical or epistemological notions like *fact* or *verification by sensory experience*.

These points are reinforced by Carnap's advocacy of the formal over the material mode of speech. The material mode involves the use of "pseudo-object sentences," which are "formulated as though they refer . . . to objects" (1937b: §74, 285). The formal mode consists of syntactic counterparts of these pseudo-object sentences. For example,

> The moon is a *thing*; five is not a thing, but a *number*

is in the material mode;

> 'Moon' is a thing-word . . .; 'five' is not a thing-word, but a number-word

is its formal mode correlate. Carnap characterizes philosophical claims in the material mode as "dangerous," "obscure," and "misleading," and asserts that "[*t*]*ranslatability into the formal mode of speech constitutes the touchstone for all philosophical sentences*" (1937b: 313).

Thus a classical positivist thesis such as

> Logical and mathematical sentences do not state anything about the facts

should be replaced by a syntactic sentence, perhaps:

> Sentences analytic in *L* are *L*-determinate.

So, for Carnap, philosophical notions like *fact* are replaced by syntactic features of languages like *L-determinacy*. Indeed, classical positivism's fundamental contrast between linguistic convention and extra-linguistic empirical fact is replaced by syntactic distinctions. As a result, there are no longer any language-independent philosophical notions. Carnap has parted company with classical positivism.

The obvious question here is: why prefer formal mode translations? Carnap states that the "*material mode of speech is not in itself erroneous*; it only readily lends itself to wrong use" (1937b: 312). Indeed, "if suitable definitions and *rules* for the material mode of speech are laid down and systematically applied, no obscurities or contradictions arise" (ibid.; my emphasis). Misuse of the material mode arises because pseudo-object terms such as 'number' "appear to designate kinds of objects, and thus make it natural to ask

questions concerning the nature of objects of these kinds" (ibid.: 310), unfortunately these questions are posed in natural languages, which are "too irregular and too complicated to be actually comprehended in a system of rules" (ibid.: 312). Carnap views natural languages behavioristically, consisting of "dispositions to certain activities, serving mainly for the purposes of communication and of co-ordination of activities among the members of a group" (1939: §2, 3), and linguistic behavior does not "determine whether the use of a certain expression is right or wrong but only how often it occurs and how often it leads to the effect intended. . . . A question of right or wrong must always refer to a system of rules" (1939: §4, 7). (See also Carnap 1937b: §2, 5; 1942: §1, 3; §5, 14; 1990: 432, final paragraph).

This diagnosis points to Carnap's attitude to the contrast between "controversies in traditional metaphysics" and "discussions in empirical science": metaphysical disputes "appeared . . . useless . . .; there seemed hardly any chance of mutual understanding, let alone of agreement, because there was not even a common criterion for deciding the controversy" (1963a: 44–45). What underlies this is the view that if indeed there is no common criterion for deciding a controversy, then it would be pointless, irrational, to continue the dispute *in the form of trying to find out who is right*. A criterion of right and wrong is supplied by a set of rules. Specifically, the abstract rules of pure syntax are projected onto linguistic behavior by "*correlative definitions*" (Carnap 1937b: 78), and this enables one to describe that behavior in the vocabulary of rational justification (see Ebbs 1997: §§57–9; Ricketts 2003: 262–4). Thus Carnap's aim, in the construction of artificial language systems, is to set out criteria that would rationalize pointless philosophical debates. This is the reason for preferring the formal mode.

Rationality would be achieved if the disputants adopted a common set of rules for adjudicating their disagreement. But this is not the only way. If rational assessment of claims is possible only on the basis of a system of rules constituting a language, it is not clear how it would be possible to assess rationally the supposed claim that that system of rules as a whole is correct or incorrect, true or false of reality. So the adoption of a system of rules is not constrained by rule-governed assessment. This is Carnap's Principle of Tolerance: "*In logic, there are no morals*. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes" (1937b: 51–2). But one could still, rationally, recommend one form of language over others, on the basis of pragmatic criteria: simplicity, convenience, safety from contradiction. So, the parties to a philosophical conflict could reconceive their opposition, not as a disagreement over the truth of a doctrine but as different recommendations about what language, what system of rules, to adopt. So reconceived, the dispute acquires a point.

None of this means that Carnap is no longer committed to empiricism. His commitment takes the form recommending certain types of languages: "it is preferable to formulate the principle of empiricism not in the form of an assertion . . . but rather in the form of a proposal or requirement [for] the language of science" (Carnap 1937a: 33). Language systems for science enable us to conceive of and to pursue science as a rule-governed rational activity. In particular, Carnap hopes to be able to display the standards governing the use of theoretical language in science as sensitive to observational evidence. But this particular verificationist form of rule-governedness is no longer essential to his opposition to metaphysics. Carnap has parted company with verificationism as a criterion of meaningfulness. What qualifies a sentence as a metaphysical pseudo-statement is simply lack of common standards of assessment. Thus Carnap writes, about the metaphysical dispute between nominalism and realism, "I cannot think of any

possible evidence that would . . . decide [it]. Therefore I feel compelled to regard [it] as a pseudo-question, until both parties to the controversy offer a common . . . indication of possible evidence regarded as relevant" (1956b: 219).

A final question that naturally arise is: What is Carnap's basis for his view of rational justification? From here, philosophical debates over the true nature of rationality loom. I suggest that Carnap's conception of rationality is also not a theoretical claim but a practical proposal, to conceive of the intractability of philosophical debate, in contrast to scientific progress, as rooted in the lack of rules for adjudicating disputes, and so to hope for philosophical progress through the replacement of traditional philosophizing by logical syntax. Carnapian pragmatism goes all the way down.

Semantics

The fundamental features of the philosophical perspective just presented remains in place when Carnap moved to semantical language-systems.

For our purposes, the main changes are the following. The primitive rules constituting linguistic frameworks are rules of designation and satisfaction, on the basis of which Carnap defines a variety of basically Tarskian truth conditions for sentences. Logico-mathematical truth is now truth in virtue of *semantic* rules of language alone, explicated by the predicate 'L-true'. The L-truth of sentences of logic is defined in terms of state-descriptions, maximal consistent sets of sentences (Carnap 1956a: §2). The analyticity predicate is defined in terms of L-truth and a class of "meaning postulates" (Carnap 1952). The crucial point is that the primitive semantic notion of reference, i.e. designation, is language-internal, since the *designata* of object language expressions are simply stated by using descriptive expressions of the meta-language translating those object expressions. So reference no more involves extra-linguistic metaphysical or epistemological notions than do the earlier syntactic concepts. This point is clear from Carnap's well-known distinction between internal and external questions (1956b). The former are to be answered by references to the rules of a linguistic framework; the latter concern the correctness of frameworks as wholes. The same line of thinking underlying Tolerance tells us that there are no rational bases for answering external questions. In particular, Carnap fails to see that ontological questions about whether a type of entities exist have any clearly defined sense when taken in a philosophical way, as independent of linguistic frameworks.

Quine Against Carnap

Classical positivism's distinctive doctrine is that logico-mathematical knowledge is non-empirical but legitimate because based on meaning or linguistic convention. Against this Quine presents three objections widely found compelling.

First, Quine argues against the thesis that the sentences of logic are true in virtue of being conventionally adopted:

> Each of these conventions is general, announcing the truth of every one of an infinity of statements conforming to a certain description; derivation of the truth of any specific statement from the general convention thus requires a logical inference, and this involves us in an infinite regress.
>
> (1976: 103)

It follows that one cannot know that the specific sentence is true by convention without knowing that the inference by which one concludes that it is an instance of the general convention is correct. This in turn requires knowledge of another logical truth, grounded on another particular instance of a general convention, and this launches the regress. So the source of knowledge of logic cannot be convention.

Second, in (1953) Quine rejects the notion of analyticity, understood as truth in virtue of meaning. One argument is that explanations of analyticity—in terms of cognitive synonymy, definition, and interchangeability *salva veritate* within a necessity operator—are circular because these other concepts are no clearer than analyticity, or are explained in terms of analyticity. What appears to be a slight variant rejects Carnap's explications of analyticity in terms of semantic rules on the ground that the notion of semantic rule remains unexplained. Against semantic rules Quine also argues that any class of truths can be labeled 'semantic rules' given an "enterprise of schooling unconversant persons in" truth condition for sentences of some language (1953: 35). But since any truth whatsoever could be used for this purpose, any truth would qualify as a semantic rule. Quine also writes that a genuine clarification has to spell out the "behavioral . . . factors relevant to analyticity" (1953: 36).

Finally, also in (1953), Quine argues against reductionism, on the basis of holism in confirmation. One can take this argument as also directed against analyticity. AS can be interpreted as an epistemic distinction, marking a difference in the bases for acceptance and rejection of sentences. The acceptance or rejection of a synthetic sentence in a language *L* is subject to evaluation as correct or incorrect on the basis of the rules of *L*. In contrast, the acceptance or rejection of an analytic sentence amounts to acceptance or rejections of the rules of *L*, i.e., of *L* itself, and so cannot be evaluated as correct or incorrect but is based on pragmatic grounds: expedience, simplicity, fruitfulness. This distinction is familiar from the argument for Tolerance, and so is plausibly ascribed to Carnap.

The holism of confirmation goes against this epistemic asymmetry. If the observational consequences of a theory conflict with actual observations, no single sentence of the theory need be singled out for rejection or disconfirmation; moreover, nothing stands in the way of rejecting the (logico-mathematical) rules according to which the consequences are drawn from the theory, or indeed the observations themselves. The choice of which sentence to revise, or whether to make any revision at all, is controlled by the same pragmatic factors that influence choice of language. Quine advances a "counter-suggestion" to AS, his web of belief model of language, also called "gradualism" (1986: 100). According to this model, each sentence of one's overall theory, or even all of science, is more or less firmly accepted, depending on the extent to which it is inferentially linked to other sentences; in response to observational evidence, the choice of which sentence to revise is shaped by how much disturbance in attitudes to other sentences a revision would cause.

Against Carnap none of these arguments is conclusive. From Carnap's deflationary perspective, the classical positivist purported thesis that conventional adoption of rules explains what makes logico-mathematical sentences true is a pseudo-statement. So Quine's regress doesn't refute any position Carnap holds. Note Carnap doesn't deny that logic is conventional. Indeed, he suggests that "principle of the conventionality of language forms" is a better name for Tolerance (1963a: 55). But this shows that for him logic is conventional only in the sense that adopting a system of logic requires no theoretical justification.

Carnap takes the circularity argument to show that natural language expressions have no systematic and precise meanings. But he fully accepts this point; one aim of linguistic frameworks is to remove such unclarities.

Quine's dismissal of semantic rules is surely one reason why Carnap found (Quine 1953) "rather puzzling" (1963b: 918). Why would one demand, as Quine does, an explanation of what makes a sentence a semantic rule? Is it that *true in virtue of semantic rules* is supposed to be a special property of some sentences that explains how we can have a type of nonempirical knowledge of what they express? So that, if we don't know what makes something a semantic rule, then we have no basis for evaluating whether truth in virtue of such rules explains the putative possession of this special epistemic property? But then Quine must seem to Carnap to have missed the fact that the purpose of frameworks, and so of semantic rules, is explication and replacement of informal philosophical pseudo-conceptions, especially that of logico-mathematical truth. Neither L-truth nor the semantic rules of a framework is meant to do any philosophical explanatory work, to capture extra-framework properties. For Carnap the choice of semantic rules is constrained by pragmatic considerations alone: for instance, effectiveness for the "enterprise of schooling unconversant persons in some language." Thus, just as in the case of the regress argument, Carnap has already rejected the position that Quine's argument seems to oppose.

Carnap tries a rational reconstruction of Quine's objection. He takes Quine to doubt that there's any clear notion to be explicated because "pre-systematic explanations of analyticity are too vague and ambiguous, . . . basically incomprehensible" (1963b: 919), and because there is no empirical evidence, in the form of linguistic behavior, for analyticity. In response, Carnap attempts to specify behavioral evidence relevant to adjudicating competing hypotheses about the meanings speakers associate with expressions (1956c: 237). This opened another line of debate with Quine, leading to Quine's notorious arguments for the indeterminacy of translation (1960). The upshot of those arguments is that behavioral evidence fails to fix even the extensions of many predicates, much less their meanings. Quine's indeterminacy arguments remain controversial. In particular, since physical theory is also not fixed by observation, it is unclear why the failure of behavioral facts to fix linguistic hypotheses should warrant the conclusion that these hypotheses are are not open to confirmation by behavioral data (see Chomsky 1969). But, even if Carnap has to concede that there is no behaviorally distinguished explicatum for analyticity, it's not clear why this invalidates his project of providing rationalizing reconceptions of philosophical disputes and scientific practice. As we have seen, Carnap doubts that the behavioral dispositions constituting the phenomena of natural languages are actually rule-governed. Carnapian language systems are not intended to state the rules speakers actually use but are tools for achieving rationality in the practice of philosophy and science.

As for Quine's anti-reductionism arguments, the holism of confirmation is no news to Carnap. Already in *Syntax* Carnap claims that an empirical "test applies, at bottom, not to a single hypothesis but to the whole system of physics . . . (Duhem, Poincaré)" (318). Moreover,

> all rules [of the physical language] may be altered as soon as it seems *expedient* to do so. This applies not only to the P[hysical]-rules but also to the *L[ogical]-rules, including those of mathematics*. In this respect, there are only differences in degree; certain rules are more difficult to renounce than others.
>
> (Ibid., emphases mine)

Finally, Carnap accepts "that 'any statement can be held true come what may'" (Carnap 1963b: 921). Carnap goes on to explain why he can take holism in stride. He distinguishes

> between two kinds of readjustment in the case of a conflict with experience, . . . a change in the language, and a mere change in . . . truth-value [of] a statement whose truth value i[s] not fixed by the rules of language. . . . A change of the first kind constitutes a radical alteration, sometimes a revolution, and it occurs only at certain historically decisive points in the development of science. On the other hand, changes of the second kind occur every minute.
>
> (1963b: 921)

In order to understand Carnap's reply, let's first turn to one of Quine's criticisms of what he calls the "linguistic doctrine of logical truth" (hereafter LD): a sentence of logic is true in virtue of the meanings of the logical constants (1963: 386). Quine considers two arguments in favor of LD, restricted to "elementary logic," i.e., first-order logic with identity. First, we are inclined to say, about proponents of systems of logic apparently consistent but contrary to our logic, that they are "merely using the familiar particles 'and', 'all', or whatever, in other than the familiar senses" (ibid.) Second, we are inclined to wonder, about so-called prelogical people who supposedly "accept certain simple self-contradictions as true," whether their affirmations haven't just been mistranslated (ibid.) In each case our reaction is supposed to be explained by LD. Quine argues that the hypothesis that the sentences of logic are "obviously true," in the sense of being very firmly accepted, explains our reactions equally well. Let's prescind from evaluating Quine's argument (see Creath 2003 for a negative assessment). I want to focus on two other points. First, Quine takes the upshot of this argument to be that LD is "empty" and has no "experimental meaning" (ibid.) Second, Carnap not only agrees with this upshot, but also says, "I am surprised that Quine deems it necessary to support this view by detailed arguments," because LD does not consist of "factual" statements but rather "proposals for certain explications" (1963b: 917).

This last point shows us how to understand Carnap's response to holism. AS, like LD, is a set of proposals for explication, that is, for rationalizing redescriptions of scientific linguistic practices in terms of certain frameworks. For this to be possible, Carnap needs no more than for the frameworks to be consistent with the holism of the (actual) practice of science. That's the point made in reply to Quine: holism of empirical testing is consistent with describing some revisions as changes in language and others as changes in theory. So the data, as it were, do not rule out AS as an explication of our confirmational practices.

All this is from Carnap's perspective. How does the situation look to Quine? Let's begin from the two planks of Quine's naturalism: "[t]here is no legitimate first philosophy, higher or firmer than physics" (1969a: 303), and, philosophy is "natural science trained upon itself and permitted free use of scientific findings" (1966: 667). We might then think that for Quine AS would be a hypothesis about our use of sentences in our practices of confirmation, on a par with any other scientific hypothesis, and gradualism a competing hypothesis. Quine should then judge the two according to the criteria—fit with observation, simplicity, clarity, expedience—that he takes to be operative in science. Does Quine then hold that gradualism comes out ahead of AS or LD? If so, he doesn't tell us why. Moreover, as we have just seen, according to Quine LD has "no

experimental meaning." So is Quine claiming that gradualism, in contrast, *has* experimental meaning? As we saw, the "data" seem consistent with but do not entail AS. Do the "data" then entail gradualism? If so, Quine gives no arguments in support. But, supposing they do, then, since they are consistent with AS, so is gradualism. So, once again, why prefer gradualism to AS?

In fact Quine doesn't take AS or LD to be on a par with gradualism. Consider his parable of Ixmann the logical positivist, "out to defend scientists against the demands of a metaphysician":

> The metaphysician argues that science presupposes metaphysical principles, or raises metaphysical problems, and that the scientists should therefore show due concern. Ixmann's answer consists in showing in detail how people . . . might speak a language quite adequate to all of our science but . . . incapable of expressing the alleged metaphysical issues. (I applaud this answer, and think it embodies the most telling component of Carnap's own anti-metaphysical representations. . . .) Now how does . . . Ixmann specify that . . . language? By telling us . . . what these [people] are to be imagined as uttering and what they are thereby to be understood to mean. Here is Carnap's familiar duality of formation rules and transformation rules (or meaning postulates). . . . But these rules are part only of Ixmann's narrative machinery, not part of what he is portraying. He is not representing his hypothetical [people] as somehow explicit on formation and transformation rules.
>
> (Quine 1963: 401)

Here it's clear that Quine does see that Carnap's linguistic frameworks are not intended as explanations of verbal behavior or scientific practice but rather as serving a practical anti-metaphysical aim. However, Quine's view of that aim isn't quite the same as Carnap's. For Carnap, as we have seen, metaphysical issues arise from a lack of rationality in colloquial language, and so will hopefully vanish through the imposition of rationalizing perspectives embodied in frameworks. The standard here is rationality; scientific practice is fully rational only as seen through a framework. For Quine, in contrast, showing that metaphysical issues need not arise in a language "adequate for all of science" is enough for dispensing with them. The standard here is the needs of science. For Quine the only use of a linguistic framework would be an "overall pattern of canonical notation" . . . "shared by all the sciences," since the "same motives that impel scientists to seek ever simpler and clearer theories . . . are motives for simplification and clarification of the broader framework" (1960: 161). Quine argues that a notation consisting of the syntax and semantics of elementary logic is sufficient for all the sciences (see 1957; 1960: chapters 3–6 provide much more detail). Thus the problem with Carnapian frameworks is that they contain more than is needed for science; all the explications of AS in these frameworks are dispensable, as far as the needs of science are concerned. Hence AS and LD are on a par with metaphysical issues; Carnap's ground (as Quine sees it) for doing without the latter "pseudo-doctrines" (Quine 1963: 390) apply no less to the former.

This is to judge Carnapian frameworks as canonical notations for science. If they are not such notations, what are they for? Perhaps here Quine is as puzzled as Carnap was by Quine's criticisms. Surely Carnap is not confused, as someone might be about Ixmann's fiction, and takes his frameworks to be hypotheses about underlying mechanisms of

actual scientists' practices? For Quine philosophizing about science is seeking a scientific account of science. From this point of view science is the production, as a result of "surface irritation and internal conditions," of language as physical phenomena, for example, "concentric air-waves," taking "the form of a torrent of discourse about" the world (Quine 1957: 1). Since the aim is a scientific account of these physical processes, Quine doesn't see the relevance of regarding science through Carnapian frameworks: He asks, "why all this creative reconstruction, all this make-believe?. . . Why not settle for psychology?" (1969b: 75).

We can now understand Quine's preference for gradualism. Gradualism pictures the doing of science in (the direction of) naturalized epistemological terms: not justification but firmness of acceptance, i.e., of assent and dissent behavior. This is of a piece with the view of science as a physical phenomenon. Neither is a scientific hypothesis, but Quine is confident of both "because we know in a general way what the world is like"; they rest on "our very understanding of the physical world, fragmentary though that understanding be" (1957: 1–2).

This proto-scientific view of science and of language is of course controversial (Chomsky 1969 outlines a classic critique; Ricketts 1982 gives a Quinean reply.) But from the perspective of our discussion what is most interesting is that if Carnap's point of view rests on a proposal for conceiving of philosophical activity and its relation to science, surely Carnap can accept Quine's view as a counter-proposal, to be evaluated on pragmatic grounds. So let me conclude this discussion with a question: does Carnap's pragmatic perspective have to count as a first philosophy, higher or firmer than science?

Final Remarks

The Carnap–Quine debate is a clash of two anti-metaphysical strategies, both opposed to classical positivism but based on opposed conceptions of language, science, and of philosophy itself. One of the ironies of the history and historiography of analytic philosophy is that Carnap is taken to be a representative of the positivism that he rejected along with Quine, and Quine's criticisms of positivism are taken to justify the reinstatement of the metaphysics that he rejected along with Carnap. With a clearer view of what is at stake we can perhaps reopen an investigation of the relation between Carnapian pragmatism and Quinean naturalism.

Related Topics

2.5 Analytic Truth
4.3 The Role of Psychology
7.4 Frege, Russell, and Wittgenstein

References

Awodey, Steve and Carus, André. (2007) "The Turning Point and the Revolution," in A. Richardson and T. Uebel eds.) *The Cambridge Companion to Logical Empiricism*, Cambridge: Cambridge University Press, pp. 165–92.

Ayer, A. J. (1936) *Language, Truth, and Logic*, London: V. Gollancz Ltd.

Bradley, F. H. (1897) *Appearance and Reality: A Metaphysical Essay*, 2nd edn., London: Macmillan.

Carnap, Rudolf. (1935) *Philosophy and Logical Syntax*, London: K. Paul, Trench, Trubner & Co., Ltd.

——. (1937a) "Testability and Meaning (Continuation)," *Philosophy of Science* 4: 1–40.
——. (1937b) *The Logical Syntax of Language*, translated by Amethe Smeaton (Countess von Zeppelin), London: K. Paul, Trench, Trubner & Co., Ltd.
——. (1939) *Foundations of Logic and Mathematics*, Chicago: University of Chicago Press.
——. (1942) *Introduction to Semantics*, Cambridge, MA: Harvard University Press.
——. (1952) "Meaning Postulates," *Philosophical Studies* 3: 65–73.
——. (1956a) *Meaning and Necessity: A Study in Semantics and Modal Logic*, 2nd edn., Chicago: University of Chicago Press.
——. (1956b) "Empiricism, Semantics, and Ontology," in *Meaning and Necessity*, pp. 205–21.
——. (1956c) "Meaning and Synonymy in Natural Languages," in *Meaning and Necessity*, pp. 233–47.
——. (1963a) "Intellectual Autobiography," in Paul Arthur Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL,: Open Court, pp. 1–84.
——. (1963b) "W. V. Quine on Logical Truth," in Paul Arthur Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court, pp. 915–22.
——. (1990) "Quine on Analyticity," in Richard Creath (ed.) *Dear Carnap, Dear Van: the Quine-Carnap Correspondence and Related Works*, Berkeley: University of California Press, pp. 427–32.
Chomsky, Noam. (1969) "Quine's Empirical Assumptions," in Donald Davidson and Jaakko Hintikka (eds.) *Words and Objections: Essays on the Work of W. V. Quine*, Dordrecht: D. Reidel, pp. 53–68.
Creath, Richard. (2003) "The Linguistic Doctrine and Conventionality," in Gary Hardcastle and Alan W. Richardson (eds.) *Logical Empiricism in North America*, Minneapolis: University of Minnesota Press, pp. 234–56.
Ebbs, Gary. (1997) *Rule-Following and Realism*, Cambridge, MA: Harvard University Press.
Hahn, Hans. (1980) "Empiricism, Mathematics and Logic," in B. McGuiness (ed.) *Empiricism, Logic, and Mathematics: Philosophical Papers of Hans Hahn*, Dordrecht: Reidel, pp. 39–42.
Hempel, Carl Gustav. (1965) *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, New York: Free Press.
Hilbert, David. (1925) "Über das Unendliche," *Mathematische Annalen* 95: 161–90.
Quine, W. V. (1953) "Two Dogmas of Empiricism," in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge: Harvard University Press, pp. 20–46.
——. (1957) "The Scope and Language of Science," *British Journal for the Philosophy of Science* 8: 1–17.
——. (1960) *Word and Object*, Cambridge, MA: MIT Press.
——. (1963) "Carnap and Logical Truth," in Paul Arthur Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court, pp. 385–406.
——. (1966) "Russell's Ontological Development," *Journal of Philosophy* 63: 657–67.
——. (1969a) "Reply to Chomsky," in Donald Davidson and Jaakko Hintikka (eds.) *Words and Objections: Essays on the Work of W. V. Quine*, Dordrecht: D. Reidel, pp. 302–11.
——. (1969b) "Epistemology Naturalized," in *Ontological Relativity and other Essays*, New York: Columbia University Press, pp. 69–90.
——. (1976) "Truth by Convention," in *The Ways of Paradox and other Essays*, revised and enlarged edn., Cambridge, MA: Harvard University Press, pp. 77–106.
——. (1986) *Philosophy of Logic*, 2nd edn., Cambridge, MA: Harvard University Press.
Ricketts, Thomas. (1982) "Rationality, Translation, and Epistemology Naturalized," *Journal of Philosophy* 79: 117–36.
——. (2003) "Languages as Calculi," in Gary Hardcastle and Alan W. Richardson (eds.) *Logical Empiricism in North America*, Minneapolis: University of Minnesota Press, pp. 257–80.
Wittgenstein, Ludwig. (1922) *Tractatus Logico-Philosophicus*, translated by C. K. Ogden, London: K. Paul, Trench, Trubner & Co., Ltd.

Further reading

For bibliographies of logical positivism, see O. Neurath, R. Carnap, C. Morris (eds.) *Foundations of the Unity of Science: Toward an International Encyclopedia of Unified Science*, 2 vols. (Chicago: University of Chicago Press, 1970); A. J. Ayer (ed.) *Logical Positivism* (New York: Free Press, 1959); R. M. Rorty (ed.) *The Linguistic Turn*, 3rd edn. (Chicago: University of Chicago Press, 1992); and Friedrich Stadler, *The Vienna Circle: Studies in the Origins, Development, and Influence of Logical Empiricism* (Vienna and New York: Springer, 2001). Sahotra Sarkar (ed.), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism*, 5 vols. (New York: Garland, 1996).

The newer scholarship on positivism in general and Carnap in particular began with Alberto Coffa, *The Semantic Tradition from Kant to Carnap: To the Vienna Station*, ed. L. Wessels (Cambridge: Cambridge University Press, 1991), Michael Friedman, *Reconsidering Logical Positivism* (Cambridge: Cambridge University Press, 1999), and papers on Carnap by Warren Goldfarb and Thomas Ricketts: Goldfarb and Ricketts, "Carnap and the Philosophy of Mathematics," in D. Bell and W. Vossenkuhl (eds.) *Science and Subjectivity* (Berlin: Akademieverlag, 1992) pp. 61–78, Ricketts, "Carnap's Principle of Tolerance, Empiricism, and Conventionalism," in Peter Clark, (ed.), *Reading Putnam* (Oxford: Blackwell, 1995), and their papers in the collections below.

Some of the best recent historical work on Carnap are by Steve Awodey and Andre Carus: Awodey and Carus, "Carnap, Completeness, and Categoricity: The Gabelbarkeitssatz of 1928," *Erkenntnis* 54 (2001): 145–72; Awodey and Carus, "Carnap's Dream: Gödel, Wittgenstein, and Logical Syntax," *Synthese* 159 (2007): 23–45; and Carus, *Carnap and Twentieth-Century Thought* (Cambridge: Cambridge University Press, 2007).

For more current scholarship see the following collections of papers: Ron Giere and Alan Richardson (eds.) *The Origins of Logical Empiricism* (Minneapolis: University of Minnesota Press, 1996); Paolo Parrini, Wesley Salmon, and Merrilee Salmon (eds.) *Logical Empiricism: Historical and Contemporary Perspectives* (Pittsburgh: University of Pittsburgh Press, 2003); Gary Hardcastle and Alan Richardson (eds.) *Logical Empiricism in North America* (Minneapolis: University of Minnesota Press, 2003); Steve Awodey and Carsten Klein (eds.) *Carnap Brought Home: The View from Jena* (Chicago: Open Court, 2004); Richard Creath and Michael Friedman (eds.) *The Cambridge Companion to Carnap* (Cambridge: Cambridge University Press, 2007); Alan Richardson and Thomas E. Uebel (eds.) *The Cambridge Companion to Logical Empiricism* (Cambridge: Cambridge University Press, 2007); and Pierre Wagner (ed.) *Carnap's* Logical Syntax of Language (New York: Palgrave Macmillan, 2009).

For careful exposition and illuminating discussion of Quine's anti-analyticity arguments see Gillian Russell, *Truth in Virtue of Meaning* (Oxford: Oxford University Press, 2008) and Cory Juhl and Eric Loomis, *Analyticity* (London: Routledge, 2009). These authors also provide sophisticated contemporary defenses of analyticity.

The newer assessment of the Quine–Carnap debate began with Howard Stein, "Was Carnap Entirely Wrong, After All?," *Synthese* 93 (1992): 275–95, and Daniel Isaacson, "Carnap and the Philosophy of Mathematics," in D. Bell and W. Vossenkuhl (eds.) *Science and Subjectivity* (Berlin: Akademieverlag, 1992), pp. 100–30. Apart from the works of Creath, Ebbs, and Ricketts cited in the References, one should also consult Alexander George, "On Washing the Fur without Wetting It: Quine, Carnap, and Analyticity," *Mind* 109 (2000): 1–24; Burton Dreben, "Quine," in Robert B. Barrett and Roger Gibson Jr. (eds.) *Perspectives on Quine* (Oxford: Blackwell, 1990); "In Mediis Rebus," *Inquiry* 37 (19944): 441–47, and "Quine on Quine," in Roger Gibson Jr (ed.) *The Cambridge Companion to Quine* (Cambridge: Cambridge University Press, 2004), pp. 287–93; and Peter Hylton, *Quine* (New York and London: Routledge, 2007.

7.6
ORDINARY LANGUAGE PHILOSOPHY

Michael Beaney

> our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our arm-chairs of an afternoon—the most favoured alternative method.
>
> (Austin, "A Plea for Excuses", p. 182)

'Ordinary language philosophy' is a somewhat misleading term for a loosely connected set of subtraditions of analytic philosophy that had their heyday in Oxford and Cambridge in the 1950s and 1960s, although their roots lie in ideas developed in the first half of the twentieth century and their influence extended well outside Britain and is still felt today. In Cambridge, the key figures were G. E. Moore (1873–1958) and Ludwig Wittgenstein (1889–1951). In Oxford, where the center of gravity of British analytic philosophy moved after the Second World War, the key figures were Gilbert Ryle (1900–76), J. L. Austin (1911–60) and P. F. Strawson (1919–2006).

1 Cambridge Ordinary Language Philosophy

Central to the development of analytic philosophy in the twentieth century was what came to be known as the 'linguistic turn.' The origins of this are disputed. Some have dated it to the use by Gottlob Frege of contextual definition in *The Foundations of Arithmetic* of 1884 (see for example Dummett 1991, pp. 111–12), or at least to the more self-conscious use of contextual definition by Bertrand Russell in his theory of descriptions, first put forward in 1905. Some have dated it even earlier, to Jeremy Bentham's use of paraphrasis in 'analysing away' talk of obligations (see for example Hacker 1996, p. 281). But there is little doubt that it received a canonical formulation in Wittgenstein's *Tractatus Logico-Philosophicus*, which was published in 1921. Here Wittgenstein wrote that "Most of the propositions and questions of philosophers arise from our failure to understand the logic of our language" (4.003). All philosophy, he went on, is 'critique of language', and he singled out Russell's theory of descriptions as showing how the grammatical form of a sentence can mislead us as to its real logical form (cf. 4.0031).

The ideas of the *Tractatus* influenced the development of both logical positivism and ordinary language philosophy. Following Russell's lead (in his introduction to the *Tractatus*), the logical positivists took Wittgenstein to be advocating an ideal logical language in which philosophical problems would not arise. In fact, however, Wittgenstein held that the propositions of our ordinary language are in perfect logical order as they are (cf. 5.5563). What was necessary was an ideal notation in which the actual logic of our language could be exhibited clearly (cf. 3.325). This belief in the fundamental adequacy of ordinary language, even if analysis is needed to reveal this, is the governing principle of ordinary language philosophy.

1.1 Wittgenstein's Later Philosophy

After the *Tractatus* was published, Wittgenstein gave up philosophy for a few years. On his return, partly prompted by discussions with members of the Vienna Circle, he started to criticize some of his earlier ideas. In particular, he rejected the assumption that there is a single logic—the propositional and predicate logic of Frege and Russell—that underlies our use of language, and the associated view that there is a single account of the way that language has meaning—that given by his so-called picture theory of language. However, Wittgenstein never repudiated the idea that our ordinary language is fundamentally adequate; what he came to stress was the different ways in which words are used and the multiplicity of our forms of language.

In his *Philosophical Investigations*, posthumously published in 1953, Wittgenstein compares words with the tools in a toolbox: "there is a hammer, pliers, a saw, a screw-driver, a rule, a glue-pot, glue, nails and screws.—The functions of words are as diverse as the functions of these objects". (§11) He introduces the idea of a 'language-game': an activity or practice in which language and actions are interwoven (cf. §7); and he gives examples to illustrate the multiplicity of language-games: from giving orders and reporting events to telling jokes, cursing and praying (§23).

Each of these language-games may have their own 'logic', or as he now prefers to call it, their own 'grammar'. There is not a single logic or grammar which fits all uses of words, no single essence to language (cf. §§ 65, 97ff.). Philosophical confusion typically arises from failing to appreciate the grammar of the relevant forms of expression. On Wittgenstein's later view, the task of philosophy is to describe the different uses of our words, to provide an '*Übersicht*' (perspicuous representation) of their grammar, in order to dissolve philosophical problems (cf. §§ 122ff.). Philosophy, he wrote, "is a battle against the bewitchment of our intelligence by means of language" (§109).

1.2 Moore

Together with Russell, Moore had rebelled against British idealism around the turn of the twentieth century, but this rebellion did not inaugurate a linguistic turn. Moore agreed with Russell in seeing the task of philosophy as the analysis of propositions. But in his early work, he regarded propositions simply as complexes of concepts, where concepts were understood as constituents of the world (1899, pp. 5, 8). Later, influenced by both Russell's theory of descriptions and Wittgenstein's work, he came to see that language can easily mislead us, and his own writing came to be characterized more and more by the meticulous—even pedantic—clarification of the meanings of ordinary propositions and the painstaking critique of philosophical claims that conflicted with those meanings.

Moore is often regarded as an ordinary language philosopher, but it might be more accurate to call him a 'common sense philosopher', as epitomized in his 1925 paper, "A Defence of Common Sense". Moore here lists a number of propositions that he claims are 'truisms', such as that he exists and has a body, as well as a further truism to the effect that other people know corresponding things. He also claims that he knows, with certainty, that these are all true. He recognizes, however, that many philosophers have disputed this, and he spends the rest of the paper attempting to show how they are wrong. A key part of his defense is the distinction he draws (influenced by Wittgenstein's *Tractatus*) between understanding the meaning of an expression and being able to give a correct analysis of its meaning (1925, p. 111). It may be that you and I disagree about what the correct analysis of the meaning of a truism is (or have no idea at all), but this does not imply that we do not understand what it means, which may be enough for us to know that it is indeed a truism.

In what sense, then, is Moore an ordinary language philosopher? An answer is provided in a paper published in 1942 entitled "Moore and Ordinary Language" by Norman Malcolm (1911–90). Malcolm met Moore and Wittgenstein in Cambridge in 1938–9 and became one of the leading practitioners of ordinary language philosophy in the United States, writing extensively on Moore's and Wittgenstein's work. Malcolm begins his paper by listing twelve philosophical statements that Moore would reject as false. One example is the statement that "We do not know for certain the truth of any statement about material things". Malcolm reports Moore's typical reply as follows: "Both of us know for *certain* that there are several chairs in this room, and how absurd it would be to suggest that we do not know it, but only believe it, and that perhaps it is not the case"! (1942, p. 347; cf. p. 351). Malcolm takes such replies to illustrate his central contention that "The essence of Moore's technique of refuting philosophical statements consists in pointing out that these statements "*go against ordinary language*" (p. 349). Saying that "I know for certain that there are several chairs in this room" (when there are) is a paradigm ordinary use of the phrase 'know for certain' (cf. pp. 354–5).

Malcolm notes that Moore might not agree with his account of Moore's method (1942, fn. 6); but in his short reply to Malcolm's paper, Moore does not repudiate that account, merely quibbling over a minor matter (1942, pp. 668–9). So we can take it that Moore accepted it as basically accurate, and with it, the clarification of his status as an ordinary language philosopher.

2 Oxford Ordinary Language Philosophy

Both Wittgenstein and Moore had a significant influence on the development of Oxford ordinary language philosophy. Ryle knew Wittgenstein well: they went on holiday together, and Wittgenstein is reported to have claimed that Ryle was one of only two people who understood his philosophy (cf. Monk 1990, pp. 275, 436). Moore's work was read and criticized by Oxford philosophers from the early 1900s onward, and Austin reputedly found Moore's style of thinking far more congenial than Wittgenstein's (cf. Hampshire 1960, p. 246; Ryle 1970, p. 283). But there was also an indigenous Oxford tradition that was highly influential, especially on Austin, although this influence was never explicitly acknowledged. This was the tradition of Oxford realism, whose central figure was John Cook Wilson (1849–1915). At various places in his posthumously published *Statement and Inference* (1926), Cook Wilson urges us to respect the distinctions embodied in ordinary language. His most explicit statement is worth quoting at length:

> The authority of language is too often forgotten in philosophy, with serious results. Distinctions made or applied in ordinary language are more likely to be right than wrong. Developed, as they have been, in what may be called the natural course of thinking, under the influence of experience and in the apprehension of particular truths, whether of everyday life or of science, they are not due to any preconceived theory. . . . the actual fact is that a philosophical distinction is *prima facie* more likely to be wrong than what is called a popular distinction, because it is based on a philosophic theory which may be wrong in its ultimate principles.
>
> (1926, II, p. 874; cf. I, pp. 34, 46, 151, 299ff., 387–8; II, p. 875)

Austin does not refer to Cook Wilson at any point in his published writings, yet Austin's famous remark, as quoted at the beginning of this chapter, is a clear echo of what Cook Wilson is saying here. That remark might be taken as the motto of ordinary language philosophy; Cook Wilson anticipated it over forty years earlier.

The view expressed here informed Cook Wilson's own methodology, in the painstaking analyses he offered of key distinctions such as that between subject and predicate, and between denotation and connotation (1926, I, Part II, chs. 5, 18). Among the Oxford realists—as there was, too, among the surviving Oxford idealists—there was also a general hostility to the logic of Frege and Russell. This was partly born of ignorance and reluctance to understand it; but it also had a philosophical motivation rooted in respect for the rich and subtle distinctions of ordinary language.

2.1 Ryle

Ryle's first significant publication was "Systematically misleading expressions" (1932). Ryle here endorses the Tractarian/Moorean distinction between understanding the meaning of an expression and knowing what its correct analysis is. The aim of philosophy, on Ryle's view, is to find the correct analysis of problematic expressions—the pursuit of which, however, does not imply that users of the expressions cannot understand them or use them intelligibly (cf. p. 43). The expressions he focuses on are those that he calls 'systematically misleading', which he characterizes as "couched in a syntactical form improper to the facts recorded and proper to facts of quite another logical form than the facts recorded" (p. 44).

The first type of such expressions that Ryle considers is what he terms 'quasi-ontological statements'. Consider, for example, the negative existential statement 'Satan does not exist', which suggests—misleadingly—that the property of non-existence is being attributed to something. If the statement is true, then there is nothing to be the subject of such an attribution. Instead, writes Ryle, it must mean something like 'nothing is both devilish and alone in being devilish' (p. 46).

Ryle considers three other types of expressions, which he calls 'quasi-Platonic expressions', which tempt us into reifying universals, 'descriptive expressions and quasi-descriptions', which are often mistakenly treated as names, and 'quasi-referential "the"-phrases', which may be wrongly taken to imply the existence of some object. In each case, Ryle argues, the expression must be reformulated to prevent the confusion from arising and to reveal the real logical form of the relevant fact. Ryle suggests, for example, that the quasi-Platonic statement 'Unpunctuality is reprehensible' should be rephrased as 'Whoever is unpunctual deserves that other people should reprove him for being

unpunctual', which makes overt reference only to the attribute of being unpunctual and not to any reified universal (p. 50).

The influence of Russell's theory of descriptions is obvious here, as is the logical atomism of Wittgenstein's *Tractatus* and Russell's writings around 1920. In his later work, however, prompted by the criticisms that were increasingly being made through the 1930s to the various forms of reductive analysis advocated in the early phase of British analytic philosophy, Ryle came to reject the assumption that there was a single logical form underlying every meaningful expression. He retained the view, however, that language contains misleading expressions that need to be clarified to avoid confusion. The paraphrase suggested above of 'Unpunctuality is reprehensible', for example, could still be offered to show that there is no need to reify universals; it is just that the paraphrase should not be taken to better reflect any 'underlying' logical form.

Ryle's later method can be described as conceptual clarification without reductive analysis—or, more simply, as connective analysis, to use Strawson's term (1992, ch. 2). In his most famous work, *The Concept of Mind*, published in 1949, Ryle talks of his aim being "to rectify the logical geography of the knowledge which we already possess" (p. 9). If we see this as connective analysis, then we can regard Ryle as continuing to insist on the distinction between understanding the meaning of an expression and knowing what its correct analysis is: one can have the former without the latter, the task of the philosopher being to provide the latter.

This idea is reflected in the distinction Ryle drew in chapter 2 between knowing how and knowing that. There are many things we may *know how* to do, such as ride a bicycle and speak a language, without being able to explain what we are doing, in other words, without *knowing that* such-and-such rules govern these activities. The task of the philosopher is to make explicit these rules, especially where confusion is likely to arise from half-hearted or inadequate attempts to articulate them. Here Ryle's key idea was that of a *category-mistake*, which occurs when things of one logical category are mistakenly assigned to a different logical category. Ryle argued that the Cartesian dogma of the 'Ghost in the Machine' was a category-mistake, classifying mental phenomena as physical events. In *The Concept of Mind*, Ryle offers a rich—if sometimes controversial—account of a whole range of concepts and distinctions embedded in our ordinary use of mental expressions, attempting to put into practice his methodology of connective analysis—charting the logical geography of our mental concepts.

2.2 Austin

Although Ryle was the more dominant figure both in Oxford and British philosophy generally in the two decades following the end of the Second World War, through the changes he effected to the study of philosophy at Oxford, his influence on appointments throughout Britain and his editorship of *Mind* from 1947 to 1972, Austin is often seen as epitomizing 'Oxford philosophy' as it came to be called from outside. Austin published little during his lifetime, but after his premature death in 1960, his papers and lectures were edited by two of his close colleagues, J. O. Urmson (1915–) and G. J. Warnock (1923–95), who made significant contributions in their own right to the development of Oxford Philosophy. Austin's *Philosophical Papers* appeared in 1961, and the two slim volumes of his lectures, *Sense and Sensibilia* and *How to Do Things with Words*, in 1962.

Austin's paper "A Plea for Excuses" (1956) provides the ideal introduction to his philosophy, both in explaining his methodology and in illustrating that methodology

in the case of the language we use in offering excuses for our conduct. Austin here makes the remark that I suggested provides the motto for ordinary language philosophy. According to Austin, "our common stock of words embodies all the distinctions men have found worth drawing" (p. 182), and during the course of the paper he shows how rich and significant are the distinctions between, for example, 'inadvertently', 'mistakenly', 'accidentally', 'absent-mindedly', 'unintentionally', 'carelessly', 'automatically', 'aimlessly' and 'purposelessly', even if at times people confuse them.

Sensitivity to the nuances of ordinary language and the multifarious things we do with language is characteristic of all of Austin's work. Austin took an interest in this both for its own sake, with an eye to developing a new type of linguistic theory, and for the dissolution it could provide of traditional philosophical problems. These two facets of his work are reflected, respectively, in the two courses of lectures, both of them given on several occasions, which were posthumously published as *How to Do Things with Words* and *Sense and Sensibilia*. *Sense and Sensibilia* might be regarded as his masterpiece: it offers a sustained critique of the doctrine, as Austin states it (in a rather Moore-like way), that "we never see or otherwise perceive (or 'sense'), or anyhow we never *directly* perceive or sense, material objects (or material things), but only sense-data (or our own ideas, impressions, sensa, sense-perceptions, percepts, &c.)" (p. 2). He takes Ayer (1940) as his main target, but the doctrine, in one form or another, was shared by many others at the time, both in and outside Oxford.

His methodology is laid down in the first lecture:

> our ordinary words are much subtler in their uses, and mark many more distinctions, than philosophers have realized . . . It is essential, here as elsewhere, to abandon old habits of *Gleichschaltung*, the deeply ingrained worship of tidy-looking dichotomies.
>
> (p. 3)

Austin attacks, in particular, the supposed dichotomy between 'sense-data' and 'material things', noting that "There is no *one* kind of thing that we 'perceive' but many *different* kinds, the number being reducible if at all by scientific investigation and not by philosophy" (p. 4). He recognizes that the so-called argument from illusion is one of the major motivations for the sense-datum doctrine but criticizes the argument as itself an illusion: there is no simple argument but rather "a mass of seductive (mainly verbal) fallacies" (p. 5). He considers the range of different cases that are often classified as 'sense-deception', including perspectival variation, refraction, mirror-images, mirages, reflections, and hallucinations; and he explains how the argument trades on a confusion between illusion and delusion (pp. 20–5). A flavour of his approach can be given by comparing the following three statements (cf. p. 36):

(1) He looks guilty.
(2) He appears guilty.
(3) He seems guilty.

While 'looks', 'appears' and 'seems' can sometimes be used interchangeably, there are clearly differences here, which the sense-datum theorist ignores in forcing all 'looks', 'appearances,' and 'seemings' into one category of phenomena that supposedly requires recognition of 'sense-data'.

How to Do Things with Words might also be read as offering a sustained critique of a prevalent philosophical doctrine: the doctrine that the primary function of a statement is to describe some state of affairs, or represent some fact, which it does either truly or falsely (cf. p. 2). Austin calls this the 'descriptive fallacy', and he introduces the term 'constative' to replace 'descriptive' when characterizing utterances that consist in making a statement (cf. pp. 3, 6). In fact, however, Austin does not directly attack this doctrine in the way that he attacked the sense-datum doctrine. His main concern is to explain what he calls 'performative'—as opposed to 'constative'—utterances. This shows, incontrovertibly, that language has many more uses than just describing states of affairs, but the clarification of performative utterances themselves takes priority. What Austin offers is a sketch of a theory of speech acts, which was to be developed further by later philosophers (see below).

A 'performative' utterance is an utterance that performs some action. For example, to say 'I promise' is itself (normally) to promise, and to say 'I bet' is itself (normally) to make a bet. Of course, various conditions have to be met for such utterances to genuinely count as performing the relevant action, and Austin discusses these—and the way in which things can go wrong or be 'infelicitous'—in some detail. He draws further distinctions to elaborate his account of performative utterances. He distinguishes the 'phonetic act' (the act of uttering certain noises), the 'phatic act' (the act of uttering certain words of a language) and the 'rhetic act' (the act of saying something meaningful). All these make up what he calls a 'locutionary act' (pp. 94–5), which he in turn distinguishes from an 'illocutionary act' and a 'perlocutionary act' (pp. 98ff.). Illocutionary acts are what we do *in* saying something, such as promising, betting, warning, informing, announcing a verdict, or criticizing. Perlocutionary acts are what we do *by* saying something, such as convincing, deterring, or surprising someone. Austin also distinguishes various types of performative utterance, depending on their illocutionary force, which he calls 'verdictives', 'exercitives', 'commissives', 'behabitives,' and 'expositives'; and the lectures conclude with a list of verbs under each heading (pp. 151–64).

Austin's invention of all this new—and as he admits himself (p. 151), somewhat rebarbative—terminology might seem surprising for an ordinary language philosopher. He is modest about whether his final classification tracks distinctions operative in ordinary language, but to the extent that it does, inventing names for them is acceptable. And even if they are contestable, this still raises our awareness of the complexities and subtleties of ordinary language. According to Austin, this is essential not only in dissolving traditional philosophical problems but also in providing the basis for a more systematic account of language. His early death prevented him from developing this in the way that had always been his longer-term aim.

2.3 Strawson

In the development of ordinary language philosophy 1950 was a key year. Ryle's *The Concept of Mind* had just been published, Austin and Strawson debated the nature of truth in the *Proceedings of the Aristotelian Society*, and Strawson's most widely read paper l, "On Referring", was published in *Mind*, announcing his own presence on the philosophical scene. I shall concentrate here on the latter.

"On Referring" is a critique of Russell's theory of descriptions. According to Russell, a sentence of the form 'The *F* is G' is to be analyzed as 'There is one and only one *F*

and whatever is *F* is G', which can be formalized in predicate logic as '$(\exists x)$ $(Fx \,\&\, (\forall y)$ $(Fy \rightarrow y = x) \,\&\, Gx)$'. To say that the present King of France is bald, for example, is to say that there is one and only one King of France and he is bald. According to Strawson, however, this theory fails to do justice to our use of definite descriptions in ordinary language. It may account for some uses but not all. For example, when I say 'The table is covered with books', I am not asserting that there is one and only one table and whatever is a table is covered with books. I am simply using the definite description 'the table' to refer to a particular table, the context making clear which table I mean; what I am saying is just that this table is covered with books. (Cf. 1950a, pp. 14–15) Moreover, where the definite description fails to refer, as in Russell's example of 'The present King of France is bald' (uttered today), the sentence is not false, as Russell thought, but truth-valueless, according to Strawson.

Strawson distinguishes between *sentences*, which can be characterized as meaningful (significant) or not, and *statements* ('assertions', or 'uses of sentences', as Strawson calls them), which are the appropriate candidates for the assignment of a truth-value; and he argues that Russell confuses the two. While 'meaning' is a function of the sentence or expression, Strawson writes, "mentioning and referring and truth or falsity, are functions of the use of the sentence or expression" (1950a, p. 9). While 'The present King of France is bald' is meaningful, according to Strawson, it fails to make, if used today, a true or false statement, just because there is a failure of reference. In his *Introduction to Logical Theory*, published two years later, Strawson distinguishes between *entailment* and *presupposition*, his point then being that in stating that the *F* is a G, one is typically *presupposing* that there is an *F*. Where there is no *F*, no statement with a truth-value is made.

The debate between Russell and Strawson has spawned a huge literature, and is familiar to all students of modern philosophy of language. I will just make two points here. First, at the heart of Strawson's critique is the distinction between sentence and statement, where making a statement is understood as a speech act. This brings him firmly on the side of Austin in stressing that a correct account of language must recognize the important role played by speech acts. Second, a continual refrain in Strawson's work is the gap between the complexity of the rules governing the use of ordinary language and the rules of formal logic, especially the predicate logic of Frege and Russell. This brings him firmly on the side of Wittgenstein in stressing the dangers of formalization in philosophy. As Strawson concludes "On Referring": "Neither Aristotelian nor Russellian rules give the exact logic of any expression of ordinary language; for ordinary language has no exact logic". (1950a, p. 27)

3 Assessment and Influence

I will end by offering a brief assessment of ordinary language philosophy and indicating something of its influence and the developments that occurred after its heyday in the 1950s and 1960s.

3.1 Assessment

In offering an assessment of ordinary language philosophy, in the space available here, I will focus on Austin, who is often taken as the paradigm ordinary language philosopher. The objections that have been made to his philosophy, and the responses that both

he and various commentators have offered, illustrate the issues that ordinary language philosophy raises. The most fundamental objection is one that Austin confronts head on in "A Plea for Excuses":

> When we examine what we should say when, what words we should use in what situations, we are looking again not *merely* at words (or 'meanings', whatever they may be) but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena.
>
> (1956, p. 182)

Austin suggests that his way of doing philosophy might be called 'linguistic phenomenology' rather than 'ordinary language', 'linguistic' or 'analytic' philosophy, which by the 1950s were the standard terms. Language is rooted in our activities and practices, so to consider how words are used in certain situations is precisely to consider what realities those situations involve. Wittgenstein was equally clear about this (see for example *Philosophical Investigations*, §§ 241, 370–3). He, too, might be called a linguistic phenomenologist, a description that helps defuse some of the misunderstandings to which ordinary language philosophy has been prone.

Austin goes on to consider two further objections. The first he calls 'the snag of Loose (or Divergent or Alternative) Usage'. Since we use language in different ways, speak loosely, and even falsely, how can we appeal to what we say, or would say in imagined situations, in either dissolving philosophical problems or developing a theory of language? Austin's answer is to deny that we disagree nearly as much as people think—and certainly not after the considered reflection that he himself encouraged in his discussions with colleagues. Divergences sometimes result, for example, simply from imagining the situations slightly differently. Far from shying away from disagreement, Austin suggests, we should pounce upon it, as the explanation of it is bound to be revealing (1956, pp. 183–4).

The second objection Austin calls 'the crux of the Last Word'. Why should ordinary language claim to have the last word on anything? A common distinction may work well for everyday practical purposes, but different distinctions may be required for scientific purposes or wherever intellectual sophistication is needed. Here Austin's answer is concessive: "ordinary language is *not* the last word: in principle it can everywhere be supplemented and improved upon and superseded. Only remember, it *is* the *first* word" (1956, p. 185). The terminology introduced in *How to Do Things with Words* provides a good example of Austin's own willingness to go beyond ordinary language in building on the first words he had so assiduously collected and divided.

Does this mean, as Maxwell and Feigl once put it in a critique of ordinary language philosophy from the perspective of logical positivism, that "many cases of putative ordinary-usage analysis are, in fact, disguised reformations" (1961, p. 194)? Austin might indeed be regarded as *proposing* that we use his refined terminology in recognizing and explaining performative utterances. But this terminology is not intended to *replace* ordinary linguistic expressions in the way that the logical positivists conceived the process of 'explication', as Carnap called it (1950, ch. 1). Austin's terminology *supplements* ordinary language, the detailed clarification of which is needed to motivate and justify that terminology.

3.2 Influence

As we have seen, Austin's work had two main purposes: to provide the basis for a systematic theory of language that recognizes the role of performative utterances and other speech acts, and to enable the dissolution of traditional philosophical problems. The theory of speech acts was developed, most notably, by Austin's student, John Searle (1932–), in *Speech Acts* (1969); and speech act theory now occupies a central place in contemporary philosophy of language. As far as the dissolution of philosophical problems is concerned, perhaps the most important of those directly influenced by Austin—as well as by Wittgenstein—is another American philosopher, Stanley Cavell (1926–), who is probably the leading advocate of ordinary language philosophy in the States. *Must we mean what we say?* (1969) and *The Claim of Reason* (1979) are his two key works.

The detailed investigations of ordinary language, encouraged by Austin and Wittgenstein, formed the basis of a research program, and there was a whole generation of philosophers brought up in the 1950s and 1960s who chose their own area of linguistic discourse to do their fieldwork (a term which Austin himself uses; 1956, p. 183). In the best work, this was always undertaken to elucidate or dissolve certain traditional philosophical problems, in just the way Austin had done in *Sense and Sensibilia* and Wittgenstein had done in *Philosophical Investigations*. To take just four prominent examples: Elizabeth Anscombe (1957) analyzed our expressions of intention, Norman Malcolm (1959) chose dreaming as his fieldwork, Anthony Kenny (1963) investigated our concepts of action, emotion, and will, and Alan White (1975, 1990) explored the jungles of our modal thinking and of our language of imagination.

H. P. Grice (1913–88) occupies a unique place in the story of ordinary language philosophy. One of his most significant early publications was a paper he wrote with Strawson defending the analytic/synthetic distinction against the Quinean critique (Grice and Strawson 1956). This placed Grice alongside the other Oxford ordinary language philosophers. He then developed his theory of conversational implicature, which offers a way of synthesizing semantic theory in the Frege–Russell mold with recognition of the role played by speech acts (1975, 1981). His main work on this occurred after he moved to Berkeley in 1967, joining Searle, so we might see this as marking the beginning of the end of the dominance of Oxford ordinary language philosophy and the development of new forms of analytic philosophy. Grice's writings have also been enormously influential on subsequent philosophy of language.

Ordinary language philosophy has also had an influence outside the analytic tradition. Perhaps the best example of this is Austin's influence on Jacques Derrida (1930–2004). Austin's attack in *Sense and Sensibilia* on "the deeply ingrained worship of tidy-looking dichotomies" (quoted above), and his recognition of the importance and complexity of what he calls 'negations and opposites' in "A Plea for Excuses" finds strong echoes in Derrida's project of deconstructing philosophical contrasts and oppositions. Derrida also had a famous debate with Searle over speech act theory, prompted by Derrida's essay "Signature Event Context" (Derrida 1972, 1977; Searle 1977).

Looking back from the twenty-first century, there was certainly a movement in British philosophy from around 1945 to 1970 for which 'ordinary language philosophy' provides a natural 'first word'. But the simple opposition between ordinary language philosophy and other forms of philosophy—whether 'ideal language philosophy', any other subtradition of analytic philosophy, or indeed so-called continental philosophy—has long since dissolved into a rich and shifting variety of subtler and smaller-scale divisions

and oppositions. This suggests that we should not seek to specify any 'essence' to ordinary language philosophy. Any opposition that we might be tempted to pose between ordinary language philosophy and other forms of philosophy needs deconstruction—a deconstruction that must involve, however, careful attention to the use of 'ordinary language' in both ordinary and philosophical discourse.

Bibliography

Anscombe, G. E. M., 1957, *Intention*, Oxford: Blackwell

Austin, J. L., 1950, "Truth", in Austin 1979, pp. 117–33; originally published in *Proc. Aris. Soc. Supp.*, 24

——, 1956, "A Plea for Excuses", in Austin 1979, pp. 175–204; originally published in *Proc. Aris. Soc.*, 1956–7

——, 1962a, *How to Do Things with Words*, eds. J. O. Urmson and M. Sbisà, Oxford: Oxford University Press

——, 1962b, *Sense and Sensibilia*, ed. G. J. Warnock, Oxford: Oxford University Press

——, 1979, *Philosophical Papers*, 3rd ed., eds. J. O. Urmson and G. J. Warnock, Oxford: Oxford University Press; 1st ed. 1961

Ayer, A. J., 1940, *The Foundations of Empirical Knowledge*, London: Macmillan

Baldwin, Thomas, 2001, *Contemporary Philosophy*, Oxford: Oxford University Press

Beaney, Michael, 2006, "Wittgenstein on Language: From Simples to Samples", in E. Lepore and B. C. Smith, eds., *The Oxford Handbook of Philosophy of Language*, Oxford: Oxford University Press, pp. 40–59

——, 2009, "Analysis", in the *Stanford Encyclopedia of Philosophy*, online at http://plato.stanford.edu/entries/analysis

——, forthcoming, *The Oxford Handbook of the History of Analytic Philosophy*, Oxford: Oxford University Press

Berlin, Isaiah et al., 1973, *Essays on J. L. Austin*, Oxford: Clarendon Press

Carnap, Rudolf, 1950, *Logical Foundations of Probability*, Chicago: University of Chicago Press

Caton, Charles E., 1963, ed., *Philosophy and Ordinary Language*, Urbana: University of Illinois Press

Cavell, Stanley, 1969, *Must We Mean What We Say?*, Cambridge: Cambridge University Press, 1976

——, 1979, *The Claim of Reason*, New York: Oxford University Press

Derrida, Jacques, 1972, "Signature Event Context", *Glyph*, 1 (1977), pp. 172–97; also in Derrida 1982, pp. 307–30; Derrida 1988, pp. 1–23; originally published in French 1972

——, 1977, "Limited Inc a b c . . .", *Glyph*, 2, pp. 162–254; also in Derrida 1988, pp. 29–110

——, 1982, *Margins of Philosophy*, translated by Alan Bass, Hemel Hempstead: Harvester; orig. publ. in French 1972

——, 1988, *Limited Inc*, Evanston: Northwestern University Press

Dummett, Michael, 1991, *Frege: Philosophy of Mathematics*, London: Duckworth

Frege, Gottlob, 1884, *Die Grundlagen der Arithmetik*, Breslau: W. Koebner, selections translated in M. Beaney, ed., *The Frege Reader*, Oxford: Blackwell, pp. 84–129

Grice, H. P., 1975, "Logic and Conversation", in Grice 1989, pp. 22–40; originally published 1975

——, 1981, "Presupposition and Conversational Implicature", in Grice 1989, pp. 269–82; originally published 1981

——, 1989, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press

Grice, H. P. and Strawson, P. F., 1956, "In Defence of a Dogma", in Grice 1989, pp. 196–212; originally published in *Phil. Rev.*, 65, pp. 141–58

Hacker, P. M. S., 1996, *Wittgenstein's Place in Twentieth-Century Analytic Philosophy*, Oxford: Blackwell

——, forthcoming, "The Linguistic Turn in Analytic Philosophy", in Beaney forthcoming

Hampshire, Stuart, 1960, "J. L. Austin", in Rorty 1967, pp. 239–47; originally published in *Proc. Aris. Soc.*, 60 (1959–60), pp. 2–14

Hanfling, Oswald, 2000, *Philosophy and Ordinary Language*, London: Routledge

Kenny, Anthony, 1963, *Action, Emotion and Will*, London: Routledge and Kegan Paul

Malcolm, Norman, 1942, "Moore and Ordinary Language", in Schilpp 1942, pp. 343–68; repr. in Rorty 1967, pp. 111–24

——, 1959, *Dreaming*, London: Routledge and Kegan Paul

Marion, Mathieu, 2000, "Oxford Realism: Knowledge and Perception", Parts I and II, *British Journal for the History of Philosophy*, 8, pp. 299–338 (Part I), 485–519 (Part II)

Maxwell, Grover and Feigl, Herbert, 1961: "Why Ordinary Language Needs Reforming", in Rorty 1967, pp. 193–200; originally published in *J. Phil.*, 58, pp. 488–98
Monk, Ray, 1990, *Ludwig Wittgenstein: The Duty of Genius*, London: Jonathan Cape
Moore, G. E., 1899, "The Nature of Judgment", in Moore 1993, pp. 1–19; originally published in *Mind*, 8, pp. 176–93
——, 1925, "A Defence of Common Sense", in Moore 1993, pp. 106–33; also in Moore 1959, pp. 32–59
——, 1942, "A Reply to my Critics", in Schilpp 1942, pp. 535–677
——, 1959, *Philosophical Papers*, London: George Allen and Unwin
——, 1993, *Selected Writings*, ed. Thomas Baldwin, London: Routledge, 1993
Passmore, John, 1966, *A Hundred Years of Philosophy*, 2nd ed., London: Penguin; 1st ed. London: Duckworth, 1957
Rorty, Richard, 1967, ed., *The Linguistic Turn*, Chicago: University of Chicago Press
Russell, Bertrand, 1905, "On Denoting", *Mind*, 14, pp. 479–93
——, 1922, "Introduction [to Wittgenstein's *Tractatus*]", in Wittgenstein 1961 [1921], pp. ix–xxii
Ryle, Gilbert, 1932, :Systematically Misleading Expressions", in Ryle 1971b, pp. 41–65; originally published in *Proceedings of the Aristotelian Society*, 32, pp. 139–70; also repr. in Rorty 1967, pp. 85–100
——, 1949, *The Concept of Mind*, London: Penguin
——, 1953, "Ordinary Language", in Ryle 1971b, pp. 314–31; orig. in *Phil. Rev.*, 52
——, 1970, "Review of 'Symposium on J. L. Austin", in Ryle 1971a, pp. 282–5
——, 1971a, *Critical Essays*, *Collected Papers*, Volume 1, Abingdon: Routledge, 2009
——, 1971b, *Collected Essays 1929–1968*, *Collected Papers*, Volume 2, Abingdon: Routledge, 2009
Schilpp, P. A., 1942, ed., *The Philosophy of G. E. Moore*, 3rd ed., La Salle, Illinois: Open Court, 1968; 1st ed. 1942, 2nd ed. 1952
Searle, John, 1969, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press
——, 1977, "Reiterating the Differences: A Reply to Derrida", *Glyph*, 1, pp. 198–208
Soames, Scott, 2003, *Philosophical Analysis in the Twentieth Century*, 2 vols., Princeton: Princeton University Press
Strawson, P. F., 1950a, "On Referring", in Strawson 1971, pp. 1–27; originally published in *Mind*, 59, pp. 320–44
——, 1950b, "Truth", in Strawson 1971, pp. 190–213; originally published in *Proc. Aris. Soc. Supp.*, 24
——, 1952, *Introduction to Logical Theory*, London: Methuen
——, 1971, *Logico-Linguistic Papers*, London: Methuen
——, 1992, *Analysis and Metaphysics: An Introduction to Philosophy*, Oxford: Oxford University Press
Stroll, Avrum, 2000, *Twentieth-Century Analytic Philosophy*, New York: Columbia University Press
Travis, Charles and Kalderon, Mark, forthcoming, "Oxford Realism", in Beaney forthcoming
Urmson, J. O., 1956, *Philosophical Analysis: Its Development between the Two World Wars*, Oxford: Oxford University Press
Warnock, G. J., 1969, *English Philosophy since 1900*, Oxford: Oxford University Press; 1st ed. 1958
——, 1989, *J. L. Austin*, London: Routledge
White, Alan R., 1975, *Modal Thinking*, Oxford: Blackwell
——, 1990, *The Language of Imagination*, Oxford: Blackwell
Wilson, John Cook, 1926, *Statement and Inference*, 2 vols., Oxford: Clarendon Press
Wittgenstein, Ludwig, 1921, *Tractatus Logico-Philosophicus*, translated by D. F. Pears and B. McGuinness, London: Routledge, 1961, 1974; originally translated by. C. K. Ogden, London: Routledge, 1922; originally published in German, 1921
——, 1953, *Philosophical Investigations*, translated by G. E. M. Anscombe, Oxford: Blackwell; 3rd ed. 1978

Further Reading

For further discussion of ordinary language philosophy and its place in the history of philosophy, see Baldwin 2001, ch. 3; Berlin *et al.* 1973; Caton 1963; Hanfling 2000; Passmore 1966, ch. 18; Rorty 1967, Part III; Soames 2003, Vol. 2, Parts 2–4; Stroll 2000, chs. 4–6; Urmson 1956, Part III; Warnock 1969, 1989. See also Hacker forthcoming on the linguistic turn; Beaney 2006 on Wittgenstein's conception of language; Marion 2000 and Travis and Kalderon forthcoming on the influence of Oxford realism on Austin and others; and Beaney 2009, §6, on analysis in analytic philosophy. Ryle discusses our talk of 'ordinary language' in Ryle 1953; and Grice reflects on Oxford philosophy in Grice 1989, chs. 10–11.

7.7
PRAGMATICS AND CONTEXT: THE DEVELOPMENT OF INTENSIONAL SEMANTICS

Jason Stanley

Introduction

In the twentieth century, logic and philosophy of language are two of the few areas of philosophy in which philosophers made indisputable progress. For example, even now, many of the foremost living ethicists present their theories as somewhat more explicit versions of the ideas of Kant, Mill, or Aristotle. In contrast, it would be patently absurd for contemporary philosophers of language or logicians to think of themselves as working in the shadow of any figure who died before the twentieth century began. Advances in these disciplines make even the most unaccomplished of its practitioners vastly more sophisticated than Kant. There were previous periods in which the problems of language and logic were studied extensively (for example the medieval period), but from the perspective of the progress made in the last 120 years, earlier work is at most a source of interesting data or occasional insight. All systematic theorizing about content that meets contemporary standards of rigor has been done subsequently.

The advances philosophy of language has made in the twentieth century are of course the result of the remarkable progress made in logic. Few other philosophical disciplines gained as much from the developments in logic as the philosophy of language. In the course of presenting the first formal system in the Begriffsscrift, Gottlob Frege developed a formal language. Subsequently, logicians provided rigorous semantics for formal languages, in order to define truth in a model, and thereby characterize logical consequence. Such rigor was required in order to enable logicians to carry out semantic proofs about formal systems in a formal system, thereby providing semantics with the same benefits as increased formalization had provided for various branches of mathematics. It was but a short step to treating natural languages as more complex versions of formal languages, and then applying to the study of natural language the techniques

developed by logicians interested in proving semantic results about formal theories. Increased formalization has yielded dividends in the philosophy of language similar to those in mathematics. It has enabled philosophers to provide better and more fruitful definitions and distinctions.

Progress in philosophy of language and logic has positively affected neighboring disciplines such as metaphysics and meta-ethics. Because of this, some philosophers have regarded philosophy of language as some kind of "first philosophy," as Descartes viewed what we would now call "epistemology." That philosophy of language has progressed significantly does not mean that it provides us with a first philosophy. One can recognize that a discipline has advanced more than others without thinking that it holds the key to all advancement. The twentieth century was the century of "linguistic philosophy," not because all or even most philosophical problems have been resolved or dissolved by appeal to language, but because areas of philosophy that involved meaning and content became immeasurably more sophisticated.

Grice and the Semantics-Pragmatics Distinction

In his extraordinarily influential paper "Logic and Conversation," Grice (1989b) set out to defend the truth-table analysis of the meaning of the natural language logical particles from the ordinary language onslaught. When Strawson spoke of the connection between antecedent and consequent that is part of the "primary use" of an English conditional statement, he spoke of the *acceptability*, *truth*, or *reasonability* of a use of a conditional statement. This suggests that Strawson did not distinguish the *truth* of an utterance from the *acceptability* of that utterance. The key to Grice's defense of the truth-table analysis of the meanings of "and," "or," and "if . . . not" is that these notions can (and often do) come apart. A given utterance can be true, even though uttering it is not acceptable because it violates conversational norms. In explaining this distinction, Grice provided the foundations for a theory of conversational norms. This theory clearly explains how an utterance may be true, though unacceptable as an assertion, due to specific facts about the conversation and its participants. Grice then used the distinction between the truth of a statement and its conversational acceptability in a defense of the thesis that the connectives of propositional logic were correct explications of their natural language counterparts. More specifically, Grice assumed that natural language logical particles have the truth-table meanings of their logical counterparts, and argued that features of the uses of these expressions that are not explicable by truth-tables are due to facts about the norms governing conversation, rather than the meanings of the words.

According to Grice, conversation is a cooperative rational activity; each conversation has a purpose. This fact about conversations imposes as a norm what Grice (1989b, p. 26) calls the *Cooperative Principle*: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged." The Cooperative Principle is the overarching principle guiding conversation. Following it imposes a number of more specific norms on conversational participants. For example, lying involves one kind of uncooperative conversational behavior, being purposely irrelevant involves another, and not being sufficiently informative is a yet a third kind of uncooperative behavior. Following the Cooperative Principle isn't always a matter of saying something true, relevant, and maximally informative; it is also a matter of how one says what one says. According to the

maxim of *manner*, one should try to list events in the order in which they occur, and to cite causes before effects.

Using these conversational principles, Grice attempts to explain many of the facts about standard use cited by Strawson without abandoning the thesis that the same truth-table analysis for logical connectives also provides the meanings of their natural language counterparts. Consider Strawson's point that "or" in ordinary language is ambiguous between an *exclusive* use ("but not both") and an *inclusive* use. Assuming that "or" unambiguously means inclusive "or" (the meaning of the logical connective for disjunction), one can explain the fact that "or" is often used exclusively by general conversational principles. Suppose Hannah uttered an instance of "P or Q," but in fact believed that both P and Q were true. Then Hannah would not be maximally informative; she would be violating Grice's conversational maxim of *quantity*. So someone believing that P and Q, to follow conversational norms should say the more informative P and Q rather than the less informative P or Q (which is compatible with the truth of only one of P and Q). So, when someone utters an instance of 'P or Q', they convey (without asserting, as part of the linguistically determined content) that they do not know that P and Q. The fact that this is part of what is conveyed by following conversational principles, rather than what is asserted as part of the linguistically determined content, can be ascertained by appeal to Grice's central criterion for distinguishing what is part of what is said (the linguistically determined asserted content) from what is merely conversationally conveyed, which is the test of *cancellability*. One can *cancel* the implication conveyed by an utterance of "P or Q" (which is that one doesn't know both P and Q) by saying "P or Q; in fact, both P and Q are true," as in "John is with Bill or he is with Frank; in fact he is with both," So, consistently with the assumption that "or" unambiguously *means* inclusive "or," one can explain why "or" is often *used* as if it meant exclusive "or."

One can use the very same kind of explanation to dissolve the sense that it is part of the meaning of a disjunctive statement that the speaker is unaware of the truth of either disjunct. If Hannah knows that John was at the party, it would be a violation of the maxim of quantity for her to assert that either John was at the party or he was at home. She would not be being maximally informative by asserting the disjunctive statement, and hence would be violating the maxim of quantity. Furthermore, the implication that the speaker is unaware of the truth of either disjunct can be canceled, as in Grice's example (1989c, pp. 44–5): "The prize is either in the garden or in the attic. I know that because I know where I put it, but I'm not going to tell you." Thus, one can explain the fact that a disjunctive statement is usually only proper if the speaker is unaware of the truth of either disjunct, without making that fact part of the conventional meanings of any words.

Grice also attempted to provide pragmatic explanations (that is, explanations from general principles governing conversation) for the divergences between the truth-table meaning for the conditional and ordinary indicative conditionals (Grice, 1989d). In attempting to account for the connection thesis, the thesis that a conditional is only assertible if the antecedent provides a ground or good reason to accept the consequent, Grice (1989d, pp. 61–2) appealed to the conversational maxims, in particular that of quantity, which directs interlocutors always to assert the strongest claim consistent with their evidence, and that of quality, which directs them to have adequate evidence for their assertions. If the indicative conditional is the material conditional, then it is true if and only if the antecedent is false or the consequent is true. If the speaker knows that

the antecedent is false, adherence to the maxim of quantity requires that the speaker simply assert the negation of the antecedent, rather than the whole conditional; *mutatis mutandis* for the truth of the consequent. So a conditional is only assertible if the speaker is unaware of the truth values of the antecedent and consequent. But the maxim of quality requires anyone who asserts a conditional to have evidence for the truth of the material conditional. Since, for the reasons just given, the evidence cannot be truth-functional (that is, the speaker's grounds cannot be knowledge of the truth-values of the antecedent or consequent), the speaker must have non-truth-functional grounds for her assertion of the material conditional, if she is adhering to the maxims of quantity and quality. So, asserting an indicative conditional, on the supposition that it has the meaning of the material conditional, requires the speaker to have non-truth-functional grounds for her assertion. More specifically, it requires the speaker to know or believe that the antecedent would be a good ground for the consequent.

As we shall see, there are a number of problems with Grice's defense of the material conditional analysis of indicative conditionals. But Grice's defense of the thesis that the meaning of "or" is exhausted by the truth-table for inclusive "or" has been widely accepted, as have a number of other Gricean explanations of use-facts. The moral of Grice's work is that the facts of linguistic use are a product of two factors, meaning and conversational norms. Failure to absorb this fact undermines many of the main theses of ordinary language philosophy.

However, there were two aspects of the ordinary language philosopher's position. The first involved emphasizing the divergences in use between the logical terms and their ordinary language counterparts. The second involved the fact that natural languages involve context-sensitive words (for example "I," "here," and "now"), and that many words only have reference relative to a context of use, and many sentences only have truth-values relative to a context of use. Since reference and truth-value are only properties of *uses* of expressions, they are inappropriate notions to use in the analysis of the linguistic meanings of expressions. In general, expression types in natural language do not have references or truth-values, only uses of them do. So employing the apparatus of semantic theory, which crucially avails itself of notions such as reference and truth, is not the right way to give a theory of meaning for natural language; the meaning of expression types is given by rules of use. Grice's response to the ordinary language philosopher only speaks to the first of these aspects of the ordinary language philosopher's position. But a response to the second aspect of the ordinary language philosopher's position was to emerge from the work of those who developed and refined intensional semantics.

The Development of Intensional Semantics: From Montague to Kaplan

Carnap's semantic theory crucially exploits the notion of a possible world in defining semantic values of expressions. Each expression has, as its primary semantic value, an intension, which is a function from a possible world to the extension of that expression at that world. In the case of sentences, the intension of a sentence is a function from possible worlds to truth-values. Carnap's semantic theory has, as its "central notion" (in Michael Dummett's sense), the notion of *truth with respect to a possible world*. The logician Richard Montague, a student of Tarski's, argued that a theory of meaning should take the more general form of *truth with respect to a context of use*, where possible worlds are but one feature of a context of use (Montague, 1974b, p. 96). Montague treated a

context of use as an *index*, a collection of semantically relevant aspects of the context of use. If the language in question contained tenses and modal operators, then the indices involved in the semantic interpretation of that language would contain times and worlds. If the language also contained the indexical terms "I" and "here," the indices would aklso have persons and places as aspects. Montague then generalized Carnap's notion of intension: instead of an intension being a function from possible worlds to extensions, an intension, for Montague, was a function from indices to extensions. For example, the intension of a sentence such as "I am tired" would be a function from indices to truth-values; it would take an index whose aspects were times, worlds, and persons to be true if and only if the person at the index was tired at the time and world of the index.

The interpretation of modal operators in Montague's system was also a generalization from their interpretation in modal semantics. In Carnap's system, the function of modal operators was to *shift* the evaluation of a proposition from one possible world to another; a modal operator took an intension, and evaluated that intension at other possible worlds. On this account, a sentence such as "possibly S" is true relative to a world w if and only if the intension of S is true in some (possibly distinct) world w'. So the function of "possible," for Carnap, is to shift the evaluation of the intension of S from w to w'; "possibly S" is true in w if and only if S is true in w' (and the function of "necessarily" is to shift the evaluation of the content of the embedded sentence to all possible worlds). In Montague's system, modal and tense operators evaluate intensions at *indices* rather than just possible worlds. On this account, a sentence such as "possibly S" is true at an index i if and only if the intension of S is true at i', where i' differs from i at most in its world feature. So rather than truth with respect to a possible world being the fundamental notion, truth with respect to an index is Montague's fundamental notion, with worlds being one element of an index. This apparatus allowed Montague to generalize the apparatus of intensional semantics to treat *context-sensitivity* in natural language without sacrificing the elegant treatment of modal and tense operators. As we shall see below, this leads to an alternative response to the challenge from ordinary language philosophy than the one developed by Grice.

Montague's contributions to the systematic study of language went well beyond generalizing intensional semantics to capturing tense and context-sensitivity. Montague's most influential papers focused on intensional constructions in natural language. In Montague (1974d), he gave an account of a number of intensional constructions other than the classic cases of propositional attitude verbs and modal and temporal contexts. For example, Montague provided a semantic analysis of *intensional transitive verbs*, such as "seek" and "worship." The difference between intensional transitive verbs and extensional transitive verbs (such as for example "kick" and "meet") is that while one cannot meet a unicorn or kick a unicorn (since there are no unicorns), one can nevertheless *seek* a unicorn. So, whereas satisfying an instance of the predicate "meeting N" requires that that there is some existent entity that one meets, satisfying an instance of the predicate "seeking N" does not require that there is some existent entity that one seeks.

Beginning with Frege and Russell intensional transitive verbs had generally been ignored in the literature, largely because of the influence of Russell's theory of descriptions, the standard method of dissolving apparent reference to nonexistent entities. Russell's theory involves providing a contextual definition of definite descriptions; meanings are assigned only to sentences containing definite descriptions rather than the definite descriptions themselves. Russell's theory helps us analyze away apparent

reference to nonexistent entities in a construction such as "John believes that the fountain of youth is in Peru," since we can apply the theory to the sentence "the fountain of youth is in Peru," and arrive at an object of John's belief, without there being a fountain of youth. In contrast, one cannot use Russell's theory to arrive at an object of seeking for a construction such as "Pizarro sought the fountain of youth," since that theory gives us no way of treating the definite description "the fountain of youth" in isolation. For this reason, Quine (1960, section 32) regimented intensional transitive verbs away in favor of propositional attitude verbs (so the intensional transitive construction "x looks for y" becomes the propositional attitude construction "x endeavors that x finds y"). The fact that analytic philosophers had not produced a successful analysis of intensional transitive verbs must be viewed as a bit of an embarrassment. The problem of intensional transitive verbs is one of the original motivations for twentieth-century discussions of content. For example, it was salient in the minds of Brentano and his students, who sought to render consistent the thesis that the characteristic feature of mental states was that they were about things, and that one could have a mental state the object of which did not exist. Montague's discussion of intensional transitive verbs was thus a watershed moment in the theory of meaning. It has subsequently given rise to a lively literature in semantics and philosophy of language on the topic (for example Partee, 1974; Zimmerman, 1993; Forbes, 2000; Richard, 2001).

Montague's semantic theory was not just distinctive for its focus on intensional constructions in natural language. Montague also returned philosophers of language and semanticists to a tradition that was lost or at the very least obscured in the kind of semantic theory favored by Tarski (and Davidson). Recall that Frege treated the traditional relation between the subject of a sentence and its predicate as that of an *argument* to a *function*. That is, Frege regarded the fundamental relationship between the semantic values of expressions in a sentence to be one of *functional application*. Though quantifiers, for Frege, had the function of binding variables within their scope, they also had determinate semantic values, namely *second-level functions*. For example "everything" denoted a function from first-level functions to truth-values. The denotation of "everything" is a function that takes any first-level function that takes every object to the true, to the true, and takes every other entity to the false. Similarly, the denotation of "something" takes to the true any first-level function that has the true for at least one value, and everything else to the false. So Frege operated with an ontology that was stratified into *types*; there were objects, then functions from objects to truth-values (first-level functions), then functions from first-level functions to truth-values (second-level functions), and on up. In Tarski's work, by contrast, no use is made of functional application as a relation between semantic values. Quantifiers are not assigned functions of various kinds; an object-language universal quantifier over objects is interpreted via the use of a meta-language quantifier over sequences. Montague's semantics returned philosophers of language to the Fregean tradition of treating semantic values as functions from arguments to values, with functional application as the primary mode of semantic composition. There are lively foundational debates between advocates of Montague's *type-theoretic* approach to semantics and advocates of the more Tarskian approach, such as James Higginbotham.

Montague's marriage of intensional semantics with type-theory was extraordinarily fruitful, and led (with the help of the work of his distinguished student Barbara Partee) to the emergence of *semantics* as a new discipline within linguistic theory. But that is not to say that the generalization of intensional semantics that was at the heart of his

program has been universally accepted; in fact, the majority of philosophers of language today regard it as incorrect. The mistake made by Montague was to think that the study of modality was a branch of *pragmatics*, the study of context-sensitivity in natural language. Recall that Montague's generalization of intensional semantics consisted of treating possible worlds as features of the more general notion of a *context of use*. He then generalized the treatment of operators as shifting the evaluation of the truth of a content from one world to the next, to shifting the evaluation of the truth of a content from one context of use (or index) to the next. It is this generalization that is widely (but not universally) regarded as an error.

The first hint that something was amiss in the assimilation of modal and temporal operators to the general study of truth relative to a context of use came from Hans Kamp's work on the temporal indexical "now." Kamp, a student of Montague's, established several theses about temporal logic. The first is there are certain natural language sentences that have readings that are most perspicuously captured via the postulation of more than just past and future tense operators. For example, consider the sentence in (1) (from Kamp, 1971, p. 231):

(1) A child was born that will become the ruler of the world.

Sentence (1) means something like "In the past, a child was born who, in the future of the present moment, becomes the ruler of the world." In order supply (1) with its natural reading, one needs to have at one's disposal an operator with the meaning of the English word "now," whose function is to evaluate its embedded content *at the present moment*. Kamp then established that a satisfactory semantics for "now" requires having two times in the Montagovian "index" that is supposed to represent a context of use. One of the times would be shifted by temporal operators such as "it was the case that" and "it will be the case that." The other time would be intuitively the time of the utterance, and would never be shifted by any operators. Its function would be to allow for the interpretation of any occurrences of "now" in the sentence. The present moment feature of the Montagovian index could not be shifted by any operators, because otherwise, in interpreting any embedded occurrences of "now" (that is, embedded inside other temporal operators), one could no longer access the present moment, and thereby successfully interpret "now".

Kamp's insights about "now" carry over directly to the modal indexical "actual." In order to provide a successful interpretation of embedded occurrences of "actual" (that is, occurrences of "actual" inside other modal operators), each Montagovian index must contain two worlds, one that would be shifted by modal operators, and the other that would be the world of utterance. Interpreting an embedded occurrence of "actual" (that is, one that occurs within the scope of other modal operators) requires keeping track of the world of utterance. For the function of the initial modal operator is to shift the evaluation of the content of the embedded sentence to another possible world, and one needs to retain the information about the actual world of utterance, in order to interpret any occurrences of "actual" within that embedded sentence. So Montague's indices each would contain two kinds of features. First, they would contain features (worlds and times) that were shifted by operators. Second, to interpret indexical operators such as "now" and "actual," the indices would contain features that were not capable of being shifted by operators but would always represent features of the actual context of use of the sentence being uttered.

Kamp's work suggests that within a single Montagovian index there are two quite different sorts of features. First, there are features that are shifted by operators, such as "necessarily" and "possible" (and the past and future tense, assuming that they are operators). Second, there are features that intuitively represent features of the actual context of use. These include the moment at which the utterance was made, and the world at which the utterance was made, which are required, respectively, to interpret indexical operators such as "now" and "actual." Furthermore, these features are not capable of being shifted by operators, or else one could not interpret embedded occurrences of indexical operators. So, for example, in evaluating the truth of the intension of say "Necessarily S" with respect to an index i, one would evaluate the intension of S at all indices i' that differed from i at most in their world index, and shared with i all the features relevant for interpreting indexical operators—that is, all those features that represent aspects of the context of use in which "Necessarily S" was uttered. This suggests that Montague's indices are not natural kinds. Each index contained two kinds of information: information relevant for interpreting modal and temporal operators, on the one hand, and information that represented features of the context of use which are relevant for interpreting indexical expressions such as "now," "actual," "I," and "here."

There were also other reasons to be suspicious of Montague's index-theoretic approach. In 1970, Robert Stalnaker pointed out (Stalnaker, 1999, pp. 36ff.)) that Montague's semantics (or as Stalnaker calls it, Montague's "Semantics-Pragmatics") did not allow for the representation of *propositions*. For Montague, there is only one semantic content of an utterance (or occurrence) of the sentence "I am tired," and that is a function from contexts of use to truth-values. If Hannah utters "I am tired," and John utters "I am tired," the only difference there is between the contents of their utterances is that one may be true and the other may be false (that is, the value of the semantic content of "I am tired" may be different, because it is being evaluated relative to distinct indices). But, as Stalnaker emphasizes, there are additional differences between their utterances. Intuitively, *what Hannah said* when she uttered "I am tired" is distinct from *what John said* when he uttered "I am tired"; they expressed different propositions. But there is no semantic value in Montague's system that represents the different propositions in question. There is just a function from indices to truth-values associated with "I am tired," and this is not the proposition expressed by either of these utterances of "I am tired" (since they express *different* propositions).

Kamp's work clearly shows the need for "double indexing." The first kind of index is required to interpret indexical expressions occurring within a sentence. The second kind of index is required to give the proper semantics for operators on content, such as "necessarily" and "possible" (and the tenses, if they are operators). But it took another student of Montague's, David Kaplan, to draw out the real moral behind the need for double indexing. In his seminal work "Demonstratives" (Kaplan, 1989), Kaplan argues that the two kinds of indices correspond to two kinds of semantic values. The first kind of index represents the dependence of semantic value upon context. The semantic content of a context-dependent sentence such as "I am tired" depends upon features of the context of use. If John is the speaker in the context of use, then "I am tired" expresses the proposition that John is tired; if Hannah is the speaker in the context of use, then it expresses the proposition that Hannah is tired. The second kind of index represents the dependence of truth of a semantic content on a *circumstance of evaluation* (such as a possible world, or a time if tenses are operators on contents), and is required to give a

satisfactory semantics for sentence-operators. A proposition may be true at one possible world but false at another.

Accordingly, expressions are associated with two kinds of semantic values, which Kaplan called *character* and *content* respectively. The character of an expression is a function from a context of use to the content of that expression relative to that context. According to Kaplan, the character of an expression is also the linguistic meaning of that expression. So, the linguistic meaning of the first-person pronoun "I" is a function from contexts of use to persons (intuitively, the speakers of those contexts). Any use of "I" has the same meaning as any other use, though a possibly distinct semantic content. Kaplan took the semantic contents of singular terms, such as proper names and indexicals such as "I" to be their referents, in Frege's sense, and he took the semantic contents of sentences relative to contexts to be propositions. Sentences operators such as "necessarily" and "possibly" shifted the world feature of the index that represented the circumstance of evaluation. The index that represented the context of use did not contain any features that were shifted by operators in the language (Kaplan, 1989, pp. 510ff.).

By dividing features of indices into contexts of use and circumstances of evaluation, Kaplan's semantic theory represents a clear advance over Montague's. It explains why only certain features are shiftable by operators and, more importantly, it gives a semantic representation of *propositions* (the values of the characters of sentences). As a result, Kaplan's distinctions have been widely adopted in philosophy of language over the last thirty years. In particular, most philosophers have come to accept that context-dependent expressions show that there are two levels of semantic value; first, linguistic meaning, and second, the content of an occurrence of an expression on an occasion. Different occurrences of an expression might have different semantic contents, despite sharing a linguistic meaning, as is so clearly the case with the first-person pronoun "I" and other indexicals.

The work done by Montague and then Kaplan allows for another kind of reply to the ordinary language philosopher's skepticism about the possibility of giving a rigorous semantics for a natural language other than the one provided by Grice. Recall that the ordinary language philosopher's skepticism arose from the conviction that truth and reference were properties of *uses* of expressions, rather than properties of expressions, and meanings were rules for using those expressions. Kaplan's semantic theory undermines these considerations. It does make perfect sense to speak of singular terms having reference, albeit *relative to a context*, and it makes perfect sense to speak of sentences having truth-values, also relative to a context. So it makes perfect sense to attribute reference and truth to expression types, once contextual relativity is factored into the semantic theory. Whereas the notion of a rule of use is vague and mystical, Kaplan's notion of the character of an expression is not only clear but set theoretically explicable in terms of fundamental semantic notions; the character of an expression is a function from a context to the reference of that expression in that context. Far from context-sensitivity being an impediment to giving a proper account of linguistic meaning in terms of reference and truth, appeal to these semantic notions allows us to give a considerably more explicit characterization of linguistic meaning than the ordinary language philosophers were capable of providing.

References

Forbes, G. (2000) "Objectual Attitudes", *Linguistics and Philosophy* 23(2): 141–183.
Grice, P. (1989a)*Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

——(1989b) "Logic and Conversation", in Grice (1989): 22–40.
——(1989c) "Further Notes on Logic and Conversation", in Grice (1989) 41–57.
——(1989d) "Indicative Conditionals", in Grice (1989): 58–85.
Kamp, H. (1971) "Formal Properties of 'Now'", *Theoria* 37(3): 227–273.
Kaplan, D. (1989) "Demonstratives" and "Afterthoughts", in J. Almog, J. Perry, and H. Wettstein, (eds.), *Themes from Kaplan*, Oxford University Press, New York.
Montague, R. (1974b)*Formal Philosophy*, New Haven: Yale University Press.
——(1974b) "Pragmatics", in Montague (1974): 95–118.
——(1974c) "On the Nature of Certain Philosophical Entities", in Montague (1974): 148–187.
——(1974d) "On the Proper Treatment of Quantification in Ordinary English", in Montague (1974): 247-270.
Partee, B. (1973) "Some Structural Analogies between Tenses and Pronouns in English", *Journal of Philosophy* 70: 601–9.
——(1974) "Opacity and Scope", in M. Munitz and P. Unger (ed.)*Semantics and Philosophy*, New York: NYU Press: 81–101.
Quine, W.V.O. (1960) *Word and Object*,Cambridge, MA: MIT Press.
Richard, Mark (2001): "Seeking a Centaur, Adoring Adonis: Intensional Transitives and Empty Terms", in P. French and H. Wettstein (eds.) *Figurative Language: Midwest Studies in Philosophy* Volume 25, Oxford: Basil Blackwell: 103–127.
Stalnaker, Robert (1999) "Pragmatics", in Stalnaker (1999): 31-46. [First published in Synthese 1970].
Zimmerman, T. E. (1993) "On the Proper Treatment of Opacity in Certain Verbs", *Natural Language Semantics* 1: 149–179.

7.8
A BRIEF HISTORY OF GENERATIVE GRAMMAR

Robert Freidin

1 The Advent of Transformational Generative Grammar

The modern inquiry into the nature, use, and origins of human language known as *generative grammar* emerged in the middle of the 20th century. It began with a conceptual shift concerning the goals of linguistic inquiry, and this first step led to further fundamental shifts in the study of language involving our understanding of language as a mental construct, the notion of grammar, and the nature of the human language faculty, of which a speaker's grammar constitutes a significant part.

The first step towards the current perspective involved a shift in focus from the external forms of a language to the grammatical system that generates it, a system that is assumed to exist in the mind of the speaker, constituting the core part of a speaker's knowledge of his or her language. From this perspective, the study of language is a subpart of several areas in philosophy, including first and foremost epistemology and the related area of philosophy of mind, in addition to philosophy of language.

The brief history of generative grammar presented here follows an essentially single line of development from the origin of transformational grammar in the 1950s to the Minimalist Program, which began in the early 1990s and continues in the present. Because of space limitations, it leaves out many details, including alternative theories that have been proposed, some still current. Thus it is not intended as a complete history of the field.

The shift in focus from external language to mental grammar took place in several steps, the first of which concerns Noam Chomsky's earliest work in linguistics, when as an undergraduate at the University of Pennsylvania in 1947 he read the proofs of Zellig Harris's *Methods in Structural Linguistics* (1951), his formal introduction to the field. Chomsky's first research project was (at Harris's suggestion) an attempt to analyze a particular language following Harris's methods. The attempt itself was a failure (see Chomsky 1975:30), but led to the first formulation of modern generative grammar—ultimately Chomsky's 1951 master's thesis *The Morphophonemics of Modern Hebrew* (henceforth MMH, published as Chomsky 1979). Thus a comparison of the two approaches is illuminating.

Harris describes *Methods* as follows:

> This volume presents methods of research used in descriptive, or more exactly, structural, linguistics. It is thus a discussion of the operations which the

> linguist may carry out in the course of his investigations, rather than a theory of the structural analyses which result from these investigations. The research methods are arranged here in the form of the successive procedures of analysis imposed by the working linguist upon his data.

Note how Harris separates the methodological operations of linguistic analysis from "a theory of structural analyses" resulting from them. *Methods* focuses solely on the former, essentially taxonomic procedures for the segmentation of a linguistic expression into its constituent parts and the classification of those elements. These operations yield what can be called a *grammar of lists*, taking the term "grammar" to be some characterization of a language. The purpose of this grammar of lists is to state all the regularities in a language, derived from an exhaustive analysis of a presumably representative corpus of utterances. Although this assumes that the finite representative corpus can be extended to the entire unbounded language, how this actually works is not addressed.

In striking contrast, *MMH* bypasses these taxonomic procedures and focuses instead on the construction of a grammar as a set of formal rules that enumerates the grammatical sentences in a language, of which the linguist's analyzed corpus is merely a special subset. For *MMH* the linguistic analysis of a language becomes "the process of converting an open set of sentences—the linguist's incomplete and in general expandable corpus—into a closed set—the set of grammatical sentences—and of characterizing this latter set in some interesting way" (*MMH*: 1). A footnote to the previous quotation addresses the unbounded character of languages: "thus the resulting grammar will contain a recursive specification of a denumerable set of sentences" (*MMH*: 67). This formulation captures the core of modern generative grammar.

In *LSLT*, written four years after MMH, Chomsky sets out three goals for research in generative grammar: constructing the grammars of individual languages, formulating a general theory of linguistic structure, and validating the results of this work. The first two goals are interdependent, as Chomsky notes (pp. 78–79).

> In constructing particular grammars, the linguist leans heavily on a preconception of linguistic structure, and any general characterization of linguistic structure must show itself adequate to the description of each natural language. The circularity is not vicious, however. The fact is simply that linguistic theory has two interdependent aspects. At any given point in its development, we can present a noncircular account, giving the general theory as an abstract formal system, and showing how each grammar is a particular example of it. Change can come in two ways—either by refining the formalism and finding new and deeper underpinnings for the general theory, or by finding out new facts about languages and simpler ways of describing them.
>
> (p. 79)

In practice the two goals are mutually reinforcing.

In the early work in generative grammar the concept of *a language* is taken to be well understood. In §56.1 of *LSLT* a language is defined simply as "a set of utterances" (p. 201). In the introduction to *LSLT*, written two decades later, the definition is given as "a set (in general infinite) of finite strings of symbols drawn from a finite "alphabet" . . . [that is] determined by general linguistic theory, in particular, by universal phonetics, which specifies the minimal elements available for any human language and provides

some conditions on their choice and combination" (p. 5). What determines membership in the set is whether a native speaker finds a given sentence acceptable (cf. Chomsky 1957: 13). However, defining language as a set of objects attempts to provide an external characterization—something whose existence is independent from individual speakers of a language, what Chomsky calls an E(xternalized)-language. (See Chomsky 1986b for a critical discussion of this notion.)

The concept of grammar is more straightforward. A grammar of a language L consists of "a system of rules that specifies the set of sentences of L and assigns to each a structural description" (*LSLT*, p. 5). The structural description of a sentence specifies all of its elements and their organization, so it follows that set of structural descriptions of the sentences of a language gives the structure of language. Beyond the linear order of lexical items, provided overtly in the phonetic representation of linguistic expressions, a structural description also involves hierarchical structure, in which lexical items group together as constituent parts. Hierarchical structure plays a major role in interpretation (for example the ambiguity of *a review of a book by two professors*), although it is entirely covert. A grammar is thus an explicit characterization of the structure of a language.

In its initial conception, a grammar consists of two distinct formal mechanisms for generating the sentences of a language: phrase structure rules and transformations. The former constructs initial syntactic structures which the latter could modify in certain ways, including the combination of pairs of initial structures—yielding complex and compound sentences. (1) gives the set of phrase structure rules for English discussed in Chomsky 1957, which for pedagogical purposes constitutes a substantial simplification of the rules developed in *LSLT*.

(1)
1. Sentence → NP + VP
2. VP → Verb + NP
3. NP → {NP_{sing}, NP_{plural}}
4. NP_{sing} → T+N+Ø
5. NP_{plural} → T+N+S
6. T → the
7. N → man, ball, etc.
8. Verb → Aux + V
9. V → hit, take, walk, read, etc.
10. Aux → C (M) (have + en) (be + ing)
11. M → will, can, may, shall, must

(Note that C in (10) stands for the element that will come to represent tense and number agreement.)

Phrase structure rules are based on a top-down analysis where a sentence is divided into it major constituent parts and then these parts are further divided into constituents, and so on until we reach the lexical items (possibly including morphemes—cf. rule 10 above) that make up the sentence. They are given as rewrite rules where the symbol to the left of the arrow is expanded (rewritten) as the string of symbols to the right of the arrow. There are two kinds of phrase structure rules in (1): those that introduce only lexical items (6–7, 9 and 11) and those that rewrite symbols as a string of constituents (1–5, 8 and 10), thereby specifying the hierarchical structure of phrases. The former introduce lexical items into derivations (lexical insertion).

Early work on phrase structure led to the conclusion that the kind of phrase structure rules under consideration are not adequate to handle certain phenomena in natural language in a simple way. For example, the discontinuous dependencies between a verbal auxiliary and the corresponding affix of the following verbal element (see the underlined forms in (2)) could not be easily expressed in a phrase structure rule format.

(2) a. Mary is helping us.
b. Bill has briefed us about the problem.

The distribution of the passive auxiliary (*be-en*) also created problems for a purely phrase structure rule account. (See Chomsky 1957, chapter 5; Postal 1964 for the original arguments, and for historical overview and evaluation, Lasnik 2000 and Freidin 2004.) These limitations were used to motivate the inclusion of transformational rules, which it was argued were better suited for capturing the abovementioned phenomena in a natural and revealing way.

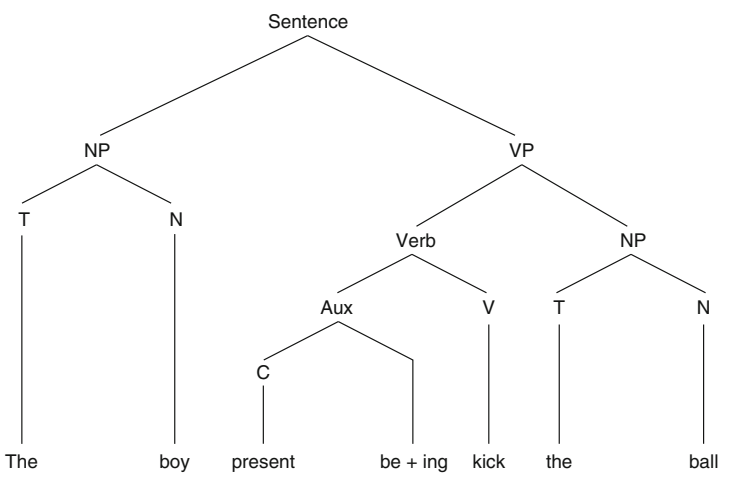
In their earliest instantiation, transformational rules operated on the set of strings generated by phrase structure rules, called a *phrase-marker*, to produce a derived phrase-marker. Consider the passive transformation from Chomsky 1957.

(3) *Passive*—optional:
Structural analysis: $NP - Aux - V - NP$
Structural change: $X_1 - X_2 - X_3 - X_4 \rightarrow X_4 - X_2 + be + en - X_3 - by + X_1$

As formulated in (3) the passive transformation identifies four consecutive constituents and rearranges them, including two insertions of lexical material (the passive auxiliary *be+en* and the passive *by*). For example, the passive transformation takes the phrase-marker underlying (4a), represented as a phrase structure tree in (5) and converts it to a phrase-marker for (4b).

(4) a. The boy is kicking the ball.
b. The ball is being kicked by the boy.

(5)



Specifying a phrase-marker for (4b) raises the problem of derived constituent structure. *LSLT* §82.1 notes two possible solutions: either specify the effect each transformation has on constituent structure in the formulation of the transformation itself or assume that general conditions apply that restrict derived constituent structure for the output of each transformation. In practice, the latter alternative is generally assumed. Moreover, it has been noted since the beginning that phrase structure rules place substantial constraints on derived constituent structure (see Chomsky 1965).

The formulation in (3) is specific to English and tied to a particular grammatical construction, the passive. In addition to the two lexical insertions, the transformation involves the transposition of two NPs. Thus this formulation involves three or four elementary operations, depending on how the transposition of NPs is analyzed (to which we will return below). The rule applies to a single phrase-marker, thus belongs to the class of *singulary transformations* in contrast to *generalized transformations*, which combine pairs of phrase-markers to produce complex and compound sentences. Furthermore, it is designated as optional—i.e., it does not have to apply when its structural description is met, in contrast to the obligatory rule that shifts verbal affixes to their appropriate positions.

From the beginning, research on transformations revealed empirical problems for the formulation and application of these rules. For example, *LSLT* §95.3 notes that while (6a) constitutes a viable sentence, the related question (6b) that might plausibly be derived from the underlying phrase-marker of (6a) via an interrogative transformation was in fact deviant.

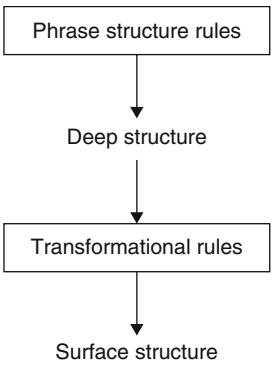
(6) a. Your interest in him seemed to me rather strange.
b. *Whom did your interest in seem to me rather strange?

Chomsky's solution in *LSLT* is to complicate the structural description of the transformation so that it does not apply to such structures. Chomsky (1964: 930–32) discusses another problem involving the interrogative transformation and this time suggests a general condition on the application of transformations, which "when appropriately formalized, might then be proposed as a hypothetical linguistic universal" (p. 931). The condition states that "if a phrase X labeled as category A is embedded within a larger phrase ZXW which bears the same label A, then no rule applying to the category A applies to X (but only to ZXW)." This condition is proposed to insure that transformations cannot apply in multiple ways to a single phrase-marker. This constitutes the first formulation of what comes to be known as the A-over-A Principle. As discussed below, it opens the way to the richest field of syntactic research: conditions on transformations.

The model of grammar that emerges has two distinct components, a set of phrase structure rules and a set of transformational rules. The output of the first yields initial underlying phrase-markers, which constitute the input to the second and yield final derived phrase-markers. Chomsky 1965 introduces a significant change to this model by introducing generalized phrase structure rules that produce embedded clauses, which allows for the elimination of generalized transformations (see §3 for further discussion). In this model initial phrase-markers are designated as *deep structures*, in contrast to the final derived phrase-marker in a derivation, which is called a *surface structure*. The model is given in (7) below.

Of the two levels of representation, surface structure is obviously the one relevant to phonological interpretation. In the case of semantic interpretation, which levels are

(7)



relevant has been the subject of considerable controversy from the 1960s on (see below for further comment). In addition, there are constraints on derived constituent structure and on the application of transformations (for example the A-over-A Principle), all of which are part of a theory of syntax.

The early work in generative grammar focuses on the form and function of grammars as the central topic for developing a general theory of linguistic structure. The other goal of constructing generative grammars of individual languages has never been fully realized. Although LSLT itself covers a substantial portion of English syntax, it was not intended as a complete survey. Matthews 1965 provides one of the earliest efforts to apply generative grammar to a non-Indo-European language. See Stockwell et al. 1973 for an attempt to survey the major syntactic structures of English. In effect the early work on transformational grammar shifted the focus of inquiry from the external forms of languages to formal grammar and the theory of linguistic structure that accounts for these forms.

2 Generative Grammar as a Theory of Knowledge

The next major development in the study of generative grammar reinforces this shift in focus. The first chapter of Chomsky 1965, written in 1958–9, interprets grammar as a model of linguistic knowledge in the mind of a speaker, consisting of a finite number of grammatical rules, including a finite lexicon. The sentences of a language are derived from the internal grammar and have no independent physical existence apart from it. In this way, language becomes a derivative notion, in contrast to grammar. Over two decades later, Chomsky 1986b: chapter 2 clarifies this in the distinction between I-language (I for internal (and in Chomsky 1995b also for individual and intentional)) vs. E-language (E for externalized). An I-language is simply the grammar in the mind of the speaker (i.e. a computational system and a lexicon that constitutes the speaker's linguistic knowledge). In contrast, an E-language is some notion of language that makes no direct reference to the mind of the speaker (for example. a set of sentences). An I-language is a finite object capable of infinite output (realizing von Humboldt's characterization of "making infinite use of finite means,"; see Chomsky 1964), in contrast to the (infinite) set of sentences that is said to comprise an E-language.

Taking a language to be a system of knowledge in the mind of the speaker raises four basic questions for the study of language (see Chomsky 1988).

(8) a. What constitutes this system of knowledge?
b. How does this system arise in the mind of a speaker?
c. How is this system put to use (in speech and writing)?
d. What mechanisms in the brain serve as the material basis for this system of knowledge and for its use?

The first question is fundamental given that answers to (8b–d) depend on the answer to (8a). In the psychological framework of (8) there is a clear separation between linguistic knowledge (*competence*) and linguistic behavior (*performance*)—see chapter one of Chomsky 1965 for discussion. While linguistic knowledge determines linguistic behavior to a large extent, the latter may be affected by "such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic)" (Chomsky 1965: 3). These effects fall under a theory of linguistic processing, ultimately leading to the ongoing debate about what effects result from the grammar vs. the processor (see Phillips 2010).

The answer to (8a) involves a lexicon and the grammatical mechanisms (i.e. phrase structure rules and transformations) that construct linguistic expressions from the items it contains. In addition the system of knowledge must include constraints on the application and output of these mechanisms—for example, the A-over-A Principle and conditions on derived constituent structure.

How such a system is acquired by a child (first language acquisition) actually involves two *how*-questions: "how is it possible?" and "how is it done step by step?" The first, which constitutes *the logical problem of language acquisition*, concerns the fact that the linguistic system acquired cannot be accounted for solely on the basis of the evidence available from the child's environment. Thus at the center of this problem lies *the poverty of the stimulus*, one of the core problems in epistemology beginning with Plato's *Meno*.

Chomsky's solution is to posit an innate—hence part of the human genome and thus universal across the species—language faculty (FL) that determines the system of knowledge acquired on the basis of the impoverished evidence provided by the environment. The resulting theory of *Universal Grammar* (UG) resurrects Descartes' theory of innate ideas in another form.

First language acquisition starts with the mind of the child in an initial state (i.e. UG) which on presentation of primary language data (PLD) changes to a steady state, the system of knowledge for the language the child comes to speak. This process is schematized in (9), where G_L is a grammar of the language in which PLD is the limited sample the child is exposed to.

(9) $UG + PLD \rightarrow G_L$

Under this framework, there are two foci for linguistic research, the content of UG and the grammars that humans acquire. UG constitutes a theory of language (as opposed to individual languages) and provides an explanatory basis for language acquisition. In contrast, G_L minus the contribution of UG merely describes the content of the linguistic knowledge acquired and therefore may raise problems for learnability if poverty of the stimulus problems remain. Ideally the contribution of UG is maximal and the remainder can be accounted for on the basis of linguistic data provided by the environment.

Perhaps the key factor involving poverty of the stimulus for language acquisition concerns a speaker's ability to distinguish deviant utterances from legitimate utterances. Thus English speakers will judge (6b) to be deviant, in contrast to the legitimate (10).

(10) To whom did your interest in Fred seem rather strange?

Given that speakers are not told to avoid sentences like (6b), their knowledge that (6b) is deviant does not apparently come from their experience. Another aspect of poverty of the stimulus arises from the fact that the PLD is finite and quite limited in contrast to the unbounded nature of the language acquired.

This unbounded character of language is a core property of language use, providing a basis for its innovative nature; thus much of what we say and hear consists of novel utterances, new to our experience and possibly to the history of the language. Language use is also free from stimulus control and moreover it is coherent and appropriate to the situation. This bundle of properties constitutes what Chomsky calls *the creative aspect of language use*, a topic first discussed in Descartes' *Discourse on Method* (part V) over three hundred years ago. (See also Chomsky 1988, 1994 for further discussion, including how this relates to the mind/body problem.) The properties of coherence and appropriateness remain a serious problem for a theory of language use (8c) in terms of both perception and production, which ultimately must involve the relation between language and thought, a relation that can only be described as mysterious given the fact that there is no formal theory of thought.

The answer to (8d) concerns a future neuroscience of language that is barely on the horizon. The major overwhelming problem for this enterprise is unification, which cannot happen when the core concepts and vocabularies of generative grammar and neuroscience remain disjoint (Poeppel and Embick 2005). Nonetheless, without doubt the brain provides the material basis for FL (including UG) as well as the I-language acquired by a speaker. In this framework both the initial state (UG) and the steady state that develops in the mind of the speaker (an I-language) constitute physical objects in the world, part of human biology. From this biolinguistic perspective, the goals and methodology of generative grammar align with those of the natural sciences in the pursuit of formal theories with a rich deductive structure (see Chomsky 1995c; Jenkins 2000).

3 From Rules to Principles and Parameters

At the outset, work on generative grammar focused primarily on the formulation of grammatical rules and their interaction in derivations. Much of this work was carried out in studies that attempted to account for the properties of constructions (for example passives and interrogatives) in various languages. However, it was clear that descriptive devices, both phrase structure rules and transformations, needed to be constrained in some principled way if the explanatory burden of the logical problem of language acquisition was to be addressed. For this reason, work on generative grammar was also concerned with placing constraints on the form and function of grammatical rules. See LSLT for the earliest proposals concerning phrase structure rules and transformations.

An important step toward constraining the descriptive power of grammatical rules was the separation of the lexicon from phrase structure rules (Chomsky 1965), in effect

separating the lexicon of a language from the mechanisms that assigned structural descriptions to linguistic expressions, what is now called the computational system of a grammar. This step eliminated substantial redundancy between the intrinsic properties of lexical items and phrase structure rules. This separation allowed for a reduction in the descriptive power of the phrase structure rule component by eliminating all context-sensitive phrase structure rules. The context-sensitive properties of lexical items (for example intransitive verbs could not occur in structures with NP objects) were handled by a separate lexical insertion operation, a substitution transformation, that matched the contextual requirements of a lexical item with the phrase-marker into which it was being inserted.

Another significant step involved the elimination of the class of generalized transformations in favor of phrase structure rules with clausal recursion. (See Chomsky 1965 for the original proposal and Freidin 1978, 1999; Lasnik 2004 for further discussion.) The application of transformations to generalized phrase-markers, now the output of phrase structure rules and lexical insertion, was constrained by the syntactic cycle, a rule-ordering principle that requires transformations to apply to the most deeply embedded cyclic domain first and then the smallest cyclic domain containing it, and so on to the root domain. Note that the syntactic cycle constitutes the first bottom-up principle of syntactic theory, in contrast to the top-down analysis imposed by phrase structure rules.

Another major development in the direction of the bottom-up analysis of phrase structure arrives in the form of X-bar theory (Chomsky 1970), where the relation between lexical categories and phrasal categories (for example N and NP) is viewed as projection of a syntactic category label of a lexical item onto the phrasal constituent formed from that lexical item and some other constituent. Note, however, that for at least another decade, the X-bar analysis continued to be formulated in terms of top-down phrase structure rules.

X-bar theory develops out of the discussion of the analysis of nominalizations like (11); see Chomsky 1968.

(11) a. the enemy's destruction of the city
b. the city's destruction by the enemy
c. the destruction of the city by the enemy

Chomsky (1970) argues against an analysis that derives (11a) from a corresponding sentence (12a) in favor of a general constraint on transformations that prohibits them from changing the syntactic categories of lexical items.

(12) a. The enemy destroyed the city.
b. The city was destroyed by the enemy.

Therefore, the examples in (11) must all be derived independently of the sentences in (12). The structural description of the NP (11a) would involve two levels of phrasal projection, originally differentiated in terms of bars, one for the first level and two for the second (given in (13) as primes).

(13) $[_{N''}$ the enemy's$]$ $[_{N'}$ destruction $[_{PP}$ of the city$]]]$

The head of this NP, *destruction*, takes *the city* as its object in the same way that the verb *destroy* takes the same NP as its object.

This parallelism between nominalization and clause extends to the passive sentence (12b) and its corresponding nominal (11b). Assuming that (11b) must be transformationally derived (as is (12b)) required a more general formulation of the "passive transformation." Chomsky 1970 splits the *Syntactic Structures* formulation (3) into two parts, a rule that postposes the underlying subject NP and another that preposes the underlying object NP—eliminating the lexical insertion operations of the passive auxiliary, which of course doesn't exist in nominals, and the passive *by*. Each part involves a single movement between two NP positions. In Chomsky 1976 such movements fall under a single rule (14).

(14) Move NP

Thus the complicated language-specific and construction-specific passive transformation of *Syntactic Structures*, which involves four distinct operations, is reduced to a single elementary operation that is neither language-specific nor construction specific. (See Freidin 1995 for a more detailed discussion of the history of the passive transformation.) Moreover, (14) generalizes to all NP movement constructions—inter-clausal raising as well as intra-clausal passive, for example subject raising cases as in (15), where *Mary* is interpreted as the subject of *like* but is pronounced as the subject of the main clause.

(15) Mary seems to like John.

The general nature of (14) makes it part of the resources of UG, rather than a language-particular rule of grammar. Later developments involve further generalizations, where movement rules are cast as single operation "Move α" (Chomsky 1981a). (See also Lasnik & Saito 1984, 1992 where movement, deletion, and insertion operations fall under a single formulation as "Affect α.")

Highly articulated structural descriptions for transformations (for example in (3)) limit the application of the elementary operations involved. Therefore when structural descriptions are minimal, as in (14), which mentions only the elementary operation and the phrase it applies to, other constraints (for example the A-over-A Principle) are necessary to stop the rule from misgenerating deviant sentences. Work on conditions on transformations is expanded in Ross's famous 1967 dissertation, which argues for replacing the A-over-A Principle with a set of constraints with a greater empirical range, proposing that certain syntactic configurations constitute islands from which movement is prohibited, primarily wh-movement. This conditions framework is further developed in Chomsky 1973 with the additions of constraints that restrict NP movement—among others including a refinement of the syntactic cycle (the Strict Cycle Condition (SCC)) and the Subjacency Condition, which subsumes some of the empirical effects of some of Ross's island constraints (the Complex NP Constraint and the Left Branch Condition) and forces movement to occur in small local increments successively cyclically rather than in a single long-distance step. The crosslinguistic generality of this conditions framework is demonstrated in Kayne (1975) on the syntax of French clitic pronouns, Quicoli 1976b on the syntax of French quantifiers, and Quicoli 1976a on the syntax of Portuguese clitic pronouns. This work shows how Chomsky's constraints on NP movement in English generalize to constructions in other languages

that do not exist in English. These analyses illustrate how linguistic variation involving different constructions actually conforms to the same general principles of grammar, in this case constraints on rule application.

Another important advance in the development of generative grammar comes shortly after with the discovery that the formulation of certain conditions may vary within a restricted range from language to language. Rizzi 1980 observes that in Italian relative clauses the relative pronoun may be extracted out of an indirect wh-question in certain constructions, but not in others—in contrast to English where indirect wh-questions constitute syntactic islands. Rizzi proposes that for Italian the Subjacency Condition is formulated with a different bounding category S' (= CP), vs. S (= TP) in English. The two values are construed as parameters for the formulation of the condition and the difference between relative clauses in Italian and English is analyzed as an instance of parametric variation. (See also Lasnik 1991 and Freidin 1992: chapters 7–8 for discussion of parameters for the principles of binding theory.) The notion of parameters is extended to other instances of crosslinguistic variation, for example the occurrence of null subject pronouns in indicative clauses (yes in Spanish; no in French and English) and the placement of interrogative phrases at the beginning of direct and indirect questions (yes in English, Spanish, French; no in Chinese and Japanese). For discussion of how parameters account for crosslinguistic variation see Chomsky 1981b; Kayne 2000, 2008; Baker 2001. See also Pollock 1989 for a parametric analysis of the verbal morphology systems of French vs. English, which has an important bearing on the syntactic analysis of clauses.

Parametric variation appears to be limited to a narrow range of choices, specified as a part of UG and thus an essential component of a theory of language design that accounts for crosslinguistic variation and complements conditions on rules of grammar, which account for crosslinguistic uniformity of language design. Taking general conditions on rules to be principles of grammar, the addition of parameters to UG creates a Principles and Parameters framework (henceforth P&P), the prevailing framework in generative grammar since 1979. This framework constitutes a significant departure from earlier work by replacing construction-specific and language-specific rules of grammar with general rules (for example Move α) that make no reference to either syntactic constructions or particular languages. In effect, grammatical operations of the computational system function as elements of UG. This perspective results in another shift in focus from grammars of particular languages to the contents and function of UG.

In another important development in the conditions framework that predates P&P, Chomsky 1973 and 1976 demonstrate how conditions on NP movement generalize to rules of semantic interpretation involving anaphor-antecedent relations. The role of semantics in generative grammar has been an important and controversial topic from the outset. A key issue is how syntactic structure supports semantic interpretation. This translates into a question about which syntactic levels of representation are relevant for interpretation. For example in passives, where the surface subject of the clause is interpreted as the logical object of the verb, the underlying structure, in which the surface subject occurs as the verbal object, is obviously relevant to interpretation. Chomsky 1957 also notes that the interpretation of quantifiers appears to depend on their linear surface order (see also Chomsky 1965: chapter 3, footnote 9) and that therefore surface structure may also play a crucial role in interpretation. (For discussion of this model and the issues concerning semantic interpretation see Chomsky 1972;

Jackendoff 1972.) An alternative model is proposed in the late 1960s in which significantly more abstract initial underlying syntactic structures are claimed to be semantic representations. (For discussion of this model of generative semantics, see McCawley 1968, Lakoff 1971; for a critique, see Chomsky 1972, especially chapter 3; and for a history of this period see Newmeyer 1986.) A third model arises in the mid-1970s in which the output of the transformational component serves as the sole level of representation for semantic interpretation. This involves a theory of movement operations in which a moved constituent leaves behind an empty constituent as a *trace* in the position from which it is moved (Chomsky 1973, Fiengo 1974). For example, the analysis of an interclausal raising construction like (15) would be (16), where the coindexed *t* marks the position of the trace of *Mary*.

(16) $[_{TP}$ Mary$_i$ $[_{VP}$ seems $[_{TP}$ t_i to $[_{VP}$ like John]]]]

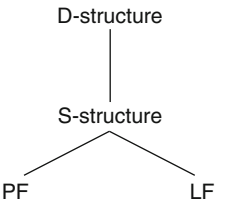
With traces, the semantic function of moved constituents (for example, the surface subject of passives) can be recovered from the output of the transformational component.

The relation between a trace and its antecedent was initially interpreted as a form of binding on a par with the relationship between a lexical anaphor (for example, a reflexive pronoun) and its antecedent. This created the option to interpret constraints on NP movement as conditions on representations (i.e., on trace binding) rather than constraints on the application of rules. Freidin 1978 demonstrates that under the former interpretation, the empirical effects of the cyclic principle (i.e., the SCC) can be derived from other independently motivated general principles, thereby eliminating the need for stipulating the SCC as a principle of grammar. This result shows the cyclic nature of syntactic derivations to be deeply embedded in the theory of grammar. The syntactic cycle remains an important topic in generative grammar—see Freidin 1999 and Lasnik 2004 for further discussion of its rich and intricate history.

The model of grammar (17) that develops under trace theory identifies four distinct levels of representation, the two levels in (7) and two levels that interface with other cognitive components: Phonetic Form (PF) with sensory-motor functions and Logical Form (LF) with conceptual-intentional functions, (17), see below.

LF is distinct from S-structure to the extent that there are additional operations that map the latter onto the former, creating covert syntactic structures necessary for semantic interpretation—for example, a rule of quantifier raising (May 1977; 1985), which creates quantifier/variable structures crucial in the interpretation of scope. Thus in this model the interface level of LF alone provides the syntactic basis for semantic interpretation.

(17)



In addition to constraints on the locality of movement operations and anaphor binding, constraints on the output of the transformational component have also been proposed, beginning with David Perlmutter's surface structure constraints in his 1968 M.I.T. dissertation (see Perlmutter 1971). This device is developed more extensively for the analysis of English in Chomsky & Lasnik 1977, which concerns an account of the distribution overt vs. phonetically null subjects in infinitival clauses among other topics. In a 1977 letter to Chomsky and Lasnik regarding their manuscript, Jean-Roger Vergnaud proposes an alternative to their surface structure filter involving notions of case and syntactic government (see Vergnaud 2008). In Chomsky 1980 (written in 1978) this proposal is formulated as the Case Filter, which prohibits the occurrence of a NP with phonetic features in a position that is not licensed for structural Case (for example nominative Case is licensed for the subject of finite clauses). This analysis is based on the universality of structural Case, regardless of whether it occurs overtly (morphologically) on all NPs (which it does not in English).

The Case Filter provides an explanation for the obligatory nature of NP movements. For example, the failure of the inter-clausal movement in (15) would yield the deviant (18), where *Mary* violates the Case Filter because it is a NP with phonetic content occurring in a syntactic position that is not licensed for Case.

(18) *It seems Mary to like John.

This analysis generalizes to NP movement in passive constructions as well under the assumption that passive predicates do not license Case on their nominal objects. In this way the syntactic behavior of Move NP is controlled by an output condition on representations rather than a constraint on operation of the rule itself.

The Case Filter has also been used account for certain facts concerning the linear order of constituents, which is merely stipulated in phrase structure rules. For example, if the adjacency of a V and its object NP is required for Case licensing, then the fact that in a VP containing an NP object and a clausal argument (CP), the linear order must be V-NP-CP and not V-CP-NP follows from the Case Filter (see Stowell 1981: chapter 3 for detailed discussion). Therefore a phrase structure rule that stipulates this order involves a significant redundancy. Note that the phrase structure rule would itself be language-specific, tied to a particular grammar, whereas the Case Filter is part of UG. Assuming this to be generally the case, Chomsky proposed in a class lecture (September 1979) that there are no phrase structure rules (see again Stowell 1981 for a detailed analysis of the redundancy between phrase structure rules and principles of UG). It is a curious fact about the history of generative grammar that although phrase structure rules were dropped from the P&P framework in 1979, no alternative formal mechanism for the construction of phrase structure was generally adopted until the proposal of an operation Merge in Chomsky 1995a (see §4 for details).

The development of Case theory (see especially Lasnik 2008) with its core principle the Case Filter was instrumental in the development of the P&P framework, leading to a modular theory of grammar in which rules and principles of grammar are organized into interacting modules. Chomsky (1981a: 135) distinguishes two classes of subsystems for the theory of grammar, as shown in (19), one for the lexicon and the rules that create from it structural descriptions of linguistic expressions at the various levels of representation (19a), and another for the principles that constrain the operation and output of the rules (19b).

(19) a. i. lexicon
ii. syntax
(1) categorial component
(2) transformational component
iii. PF-component
iv. LF-component
b. i. bounding theory
ii. government theory
iii. θ-theory
iv. binding theory
v. Case theory
vi. control theory

Bounding theory is concerned with the locality of movement operations for which the Subjacency Condition is a core principle. θ-theory covers predicate/argument structure, governed by the θ-Criterion (Chomsky 1981a), which requires in part that every argument in a linguistic expression must be assigned an argument function by a predicate. Government theory employs the concept of syntactic government to further constrain movement operations by requiring that in addition to satisfying the Subjacency Condition, empty categories such as traces must be properly governed (the Empty Category Principle (ECP)). Chomsky 1981a attempts to unify the subcomponents of (17b) by formulating many of the basic principles in terms of syntactic government, yielding the so-called Government and Binding (or GB) Theory. See also Chomsky 1986a for an attempt to reformulate the Subjacency Condition in terms of government.

Although a decade of research in GB theory produced an explosion of new data from a variety of languages, inaugurating a rebirth of comparative grammar, the concept of syntactic government—the heart of the theory—proved to be problematic and was eventually dropped in the early 1990s (see Chomsky 1993). The main weakness resided in the complex formulations of the relations "govern" and "properly govern." Various formulations of the ECP were so complex that they mirrored the complexity of the data they were supposed to account for. Nonetheless, GB yielded major developments in the P&P framework, which remains a major framework in generative grammar today.

4 Economy and the Minimalist Program

Notions of simplicity and economy for syntactic theory had been discussed in the first formulation of generative grammar (see Chomsky 1979 and for discussion, Freidin 1995; Freidin & Lasnik 2011). In the mid-1980s the notion of economy resurfaces explicitly in the form of a principle of Full Interpretation (henceforth FI, first proposed in Chomsky 1986b), which requires that every element in PF and LF representations receive an appropriate interpretation. In 1991 Chomsky construes FI as a prohibition against superfluous symbols in the two interface representations. An argument (NP or clause) in a syntactic representation that is not assigned a semantic function is uninterpretable, hence superfluous and therefore in violation of FI. Thus FI subsumes part of the θ-Criterion.

The inclusion of FI as a part of UG has other far-reaching consequences and therefore is more than just a restatement of part of the θ-Criterion. FI also rules out sentences containing any superfluous lexical items—verbs, prepositions, and adjectives as well

as nouns. Under the assumption that a quantifier in natural language to be interpretable must bind a variable, FI requires some form of quantifier raising, where the trace of the raised quantifier functions as a variable (May 1985). This generalizes to wh-movement in questions where the wh-phrase functions at LF as a quasi-quantifier (Chomsky 1976) and more generally to other syntactic operators like relative pronouns (Chomsky 1977).

FI also determines the nature of syntactic derivations. Given that phonetic features are irrelevant for the interpretation of LF and correspondingly that semantic features are likewise irrelevant for the interpretation of PF, FI requires that there be a point in a derivation where the two sets of features are separated, called *Spell-Out* (Chomsky 1993). Spell-Out, in contrast to S-structure in (17), is not a level of representation because there are no conditions that define it, nor could there be if derivations involve multiple points of Spell-Out as proposed in Uriagereka (1999) and Chomsky (2001; 2008).

In addition to the single economy condition on representations, FI, there are also economy conditions on derivations concerning the nature and function of movement operations. One involves the "last resort" character of the major movement operations for *wh*-phrases and non-*wh*-NPs, named Last Resort (Chomsky 1986b, 1995b; Lasnik 1995). In the case of *wh*-phrases movement is required by FI to create operator-variable constructions at LF. Thus movement happens because it must. The same holds for non-*wh*-NP movement, as in simple passives (for example (4b) and (12b)) and raising constructions (for example (15)), as required by the Case Filter, possibly FI under the appropriate analysis. For example, Last Resort would block interclausal NP movement out of finite clause, as in the deviant (20a) in contrast to the legitimate (20b), where the subordinate clause subject does not move.

(20) a. *[$_{TP}$ Mary$_i$ [$_{VP}$ seems [$_{TP}$ t_i [$_{VP}$ likes John]]]]
b. It seems Mary likes John.

Since *Mary* is licensed for Case as the subject of the subordinate clause, the NP has no reason to move. This analysis generalizes to nonsubject NPs as illustrated in (21), where *John* is interpreted as the object of *like*.

(21) a. *John seems Mary to like.
b. *John seems Mary likes.

Another economy condition on derivations concerns the fact that syntactic movement seems to involve shortest steps (as in successive cyclic movement), a requirement that has a "least effort" flavor. One formulation, the Minimal Link Condition (MLC), constrains the links of a chain created by traces and their lexical antecedent (Chomsky 1995a, 1995b, 2001, 2008; Boeckx 2003).

The reconsideration of the various modules of GB theory in terms of notions of economy leads not only to a simplification of some parts of GB but, more importantly, to the formulation of a minimalist program (henceforth MP) for linguistic theory (Chomsky 1993, 1995b; for a discussion of the roots of the MP, see Freidin & Lasnik 2011, and for discussion of evolution of the MP within the generative enterprise as a part of the scientific enterprise, see Freidin & Vergnaud 2001). The MP attempts to address two fundamental questions (22), first articulated in Chomsky 1998.

(22) a. To what extent is the computational system for human language optimal?
b. To what extent is human language a perfect system?

A major task for the MP is to provide explicit interpretations for the terms "optimal" and "perfect." Note also that the two questions are interrelated in that an optimal computational system for human language would almost certainly be a prerequisite if human language were actually a perfect system.

The P&P framework, which serves as a foundation for the MP, is itself based on several fundamental assumptions that suggest that the answers to both questions could be "to a significant extent." Among the core assumptions is the definition of a language (i.e. an I-language) as consisting of a lexicon and a computational system, where the latter is largely determined by UG—i.e. the grammatical operations that construct structural representations of linguistic expressions and the constraints on their operation and output. Given that knowledge of a lexicon also involves poverty of stimulus issues, it appears that some part of the lexicon is, like syntactic structure, determined by properties of UG (see Chomsky 2000: chapter 2), possibly including parameters (see Borer 1984). From this perspective, given UG a selection S of parameters determines a language. Possibly the sole task of language acquisition is acquiring a lexicon, which results in fixing the value of parameters. If the locus of linguistic variation rests solely in the lexicon (for example) the phonetic labels of lexical items (what Chomsky calls "Saussurean arbitariness") and parameters, then there may be no variation in overt syntax or the LF component. This leads to a striking follow-up conjecture that "there is only one computational system and one lexicon" (Chomsky 1993: 3), apart from idiosyncratic lexical properties. This conjecture remains controversial.

The initial formulation of the MP in Chomsky 1993 articulates some additional assumptions that are unique to the MP. Foremost among them, and the one that essentially underpins the rest, is that the theory of grammar must meet a criterion of conceptual necessity. In essence the assumption constitutes a version of Ockham's razor (the law of parsimony), a methodological consideration that has informed modern science (including modern generative grammar) for six centuries. However, the MP applies this more aggressively and rigorously than previous work in the field, yielding a substantive hypothesis that the only conceptually necessary linguistic levels of representation are the two interface levels PF and LF. Thus the MP discards D-structure and S-structure, the two levels of representation internal to derivations, and attempts to reanalyze the empirical evidence that was employed to motivate them. Without internal levels, the only output conditions possible apply to the two interface levels. Each linguistic expression is a realization of interface conditions. Furthermore, if the output conditions imposed on LF and PF representations are determined by the cognitive components that interface with them, then language would constitute "an optimal way to link sound and meaning"—designated as the Strong Minimalist Thesis (SMT). "If SMT held fully, which no one expects, UG would be restricted to properties imposed by interface conditions. A primary task of the MP is to clarify the notions that enter into SMT and to determine how closely the ideal can be approached" (Chomsky 2008: 135).

The SMT, if correct, is one aspect of language design that points to the optimal nature of the computational system. Another concerns what grammatical operations the system contains. Under P&P, the transformational component has been reduced to a small set of elementary operations, including "Move α" a deletion operation that, in addition to

eliminating non-trivial chains at PF (see discussion of (28) below), would account for ellipsis phenomena as illustrated in (23) where the phonetic features of the VP *written about Descartes* are deleted in the second conjunct (as indicated by the strikethrough).

(23) Jane has ~~written about Descartes~~ and Bill has too.

This formulation of the transformational component also supports the hypothesis for the optimal nature of the computational system.

With the elimination of phrase structure rules (see above), the question of what grammatical mechanisms replace these rules becomes crucial for the MP. The answer is provided in Chomsky 1995a, formulated as the operation Merge, which joins two syntactic objects (lexical items and/or objects formed from lexical items) and projects the syntactic label (the syntactic category feature) of one of them as the label of the new syntactic object created. Merge is a recursive operation (i.e. it applies to its own output) that builds phrase structure bottom-up. Thus consider the derivation under Merge for (24).

(24) Our group will leave early.

The two lexical items *leave* and *early* merge to form a VP, which in turn is merged with the modal *will* to form a T(ense) Phrase (TP). Independently, *our* and *group* are merged to form a NP. Finally, this NP and T phrase *will leave* early are merged to form a TP—i.e. the syntactic category T is projected as the label of clause formed. This last step merges two independent phrases in essentially the same way that generalized transformations operated in the earliest transformational grammars. In effect, a general transformational operation replaces a set of less general and more complicated phrase structure rules, including those rules that are reduced to their most general schema under X-bar theory. What this shows is that a pure transformational grammar (i.e. without phrase structure rules) constitutes a best solution for the construction of phrase structure. It is a surprising development given the abandonment of generalized transformations in favor of phrase structure rules that occurred in 1965.

If a simple operation like Merge constitutes the only grammatical mechanism for building basic phrase structure, then it also supports the optimality hypothesis the MP seeks to establish. Note too that constraining Merge also contributes to the hypothesis. While Merge applies bottom-up, unconstrained it could presumably insert a syntactic object inside an existing syntactic object. Such derivations are prohibited by the No Tampering Condition (NTC) of Chomsky 2007, 2008, formulated in (25).

(25) Merge of X and Y leaves the two S[yntactic] O[bjects] unchanged.
(Chomsky 2008:138)

The NTC constitutes a general constraint on derived constituent structure and at the same time enforces the cyclic application of Merge by allowing only merger at the root for all syntactic objects. Moreover, the NTC provides a natural constraint on efficient computation, another contribution to the optimality hypothesis.

If the NTC imposes cyclic computation in syntax, then it applies not only to the single generalized transformation Merge but also to the single singulary transformation Move α. This holds without further stipulations if Move α is actually a form of Merge, referred to as *Internal Merge* (IM) to distinguish it from *External Merge* (EM),

which operates on two independent syntactic objects. Thus consider the application of Move α that is involved in generating the passive construction (26).

(26) The review was written by two professors.

Given that the NP *the review* is first merged as the object of *written*, Merge constructs (27) prior to the application of the movement operation.

(27) [$_{T}$ was [$_{VP}$ [$_{V}$ written [$_{NP}$ the review]] [$_{PP}$ by [$_{NP}$ two professors]]]

To generate (26) from (27), all that is required is that a copy of the NP *the review* be merged with (27), yielding (28).

(28) [$_{TP}$ [$_{NP}$ the review] [$_{T}$ was [$_{VP}$ [$_{V}$ written [$_{NP}$ the review]] [$_{PP}$ by [$_{NP}$ two professors]]]]

At PF the copy of *the review* in VP is deleted, thus linearizing the nontrivial chain in (28) (see Nunes 2004 for details). Under this analysis, there is a single operation Merge that accounts for all of phrase structure, including the contribution of singulary transformations. This reduction of phrase structure rules and movement transformations to a single recursive operation provides a solid foundation for the optimality hypothesis of the MP.

It is worth noting that the NTC in conjunction with this analysis of Merge makes the generation of cannonical D-structure impossible. Consider the derivation of (29).

(29) They claimed that the review was written by two professors.

The NTC requires that the copy of *the review* that occurs as the subject of the subordinate clause be merged before the complementizer *that* is merged with that clause. Thus there is no way of segregating IM operations from EM operations and applying the former as a block after the latter have applied. In this way, an optimal formulation of the computational system supports the minimalist assumption that the language faculty involves only interface representations, providing some justification for the rigorous application of a criterion of conceptual necessity.

Given the plausibility of the optimality hypothesis, there is reason to expect that the hypothesis that language constitutes a perfect system can be explicated in some substantive way. One of the earliest proposals is that the computational system utilizes only the features found in lexical items and therefore "interface levels consist of nothing more than arrangements of lexical features," what is called the Inclusiveness Condition (IC) (Chomsky 1995b: 225 –but see page 228 and footnote 10 for comments on how inclusiveness might be violated by PF representations). It follows from this condition that interface representations will not contain bar levels on syntactic categories as in X-bar theory (cf. (13)), nor will labels distinguish maximal phrasal projections (XP) from heads (X) or their intermediate phrasal projections. In addition the IC prohibits the use of indices in representations (cf. (16)), which significantly affects the formulation of binding theory (Chomsky 1993; Freidin 1997), and also the postulation of traces as special empty categories (see again (16)). In effect, the IC enforces the copy theory of IM, where non-trivial chains consist of multiple copies of a phrase. Note that the IC is

actually the null hypothesis for how the lexicon interacts with the computational system to produce structural representations of linguistic expressions. Likewise, the NTC also functions as a null hypothesis on the output of Merge.

A computational system that contains Merge (i.e. both EM and IM) governed by the NTC and the Inclusiveness Condition constitutes a highly constrained theory of derived constituent structure, ultimately interface representations. To the extent that the functioning of this computational system is largely determined by interface conditions imposed by the cognitive components that interact with the language faculty—i.e. the components which operate on the interface representations PF and LF and thereby make language usable, it could be that I-language is a perfect solution to interface conditions. However, the locality conditions on IM, going back to the A-over-A Principle and Ross's island constraints, remain to be accounted for. From an economy perspective it appears that IM involves some requirement for "shortest move" which might translate into a "minimal link" constraint on chains (see references on p. 909). If this constraint applies to derivations, specifically the application of IM, then it is part of the computational system, necessarily a principle of UG. Alternatively, this requirement could be formulated as a condition on representations (see Freidin 1978 for a formulation of Subjacency as a condition on trace binding), in which case it would be possible for the locality of IM, which requires successive cyclic movement, to follow from interface conditions. Exactly which analysis is correct remains to be determined.

From the discussion above, it should be clear that the development of the MP has resulted in shrinking the content of UG. To the extent that interface conditions are legibility requirements imposed on the sole interface representations PF and LF by the cognitive components that connect with linguistic representations, they are not part of UG. This leaves the operations Merge and Delete, and conditions on their application—including the NTC, the IC, and some version of the MLC, along with a condition on the recoverability of deletions (Chomsky 1965)—as parts of UG.

This development has a significant impact on the potential answer to the prime biolinguistic question: Why is it that humans acquire natural languages and other species do not? Because the shrinking of UG highlights the computational core of language—i.e. Merge, the recursive procedure that accounts for linguistic structure, it could be that recursion is the answer to the biolinguistic question (see Hauser et al. 2002). This answer undoubtedly bears on the question of the evolutionary origins of language as a system of knowledge, however it remains controversial—see Pinker & Jackendoff (2005), Fitch et al. 2005, Jackendoff & Pinker 2005, and Larson et al. 2010 for discussion.

The shrinking of UG also raises the question of how much (or how little) needs to be postulated in UG to account for language design. This question arises in a more highly articulated framework than was available in earlier work where the dichotomy of nurture vs. nature in language acquisition was generally conceived of as a two-way split between the external language environment (PLD) and the innate—hence internal—UG that largely determined the grammar acquired. More recent discussions (see Chomsky 2005, 2007) have raised the possibility that a third factor—namely, principles not specific to the language faculty—could also play a role. Although the possibility of third-factor explanations for language is briefly mentioned in Chomsky 1965, it is only with the development of the MP, especially its concern for principles of efficient computation, that third-factor accounts become plausible. Consider what Chomsky wrote:

> there is surely-no reason today for taking seriously a position that attributes a complex human achievement entirely to months (or at most years) of experience, rather than to millions of years of evolution or to principles of neural organization that may be even more deeply grounded in physical law—a position that would, furthermore, yield the conclusion that man is, apparently, unique among animals in the way in which he acquires knowledge.
>
> (p. 59)

Principles of neural organization would not be specific to language, hence a third factor. If such principles follow from physical law applied to neurons, then an account of language design based on them would almost certainly involve optimal computation in a perfect system.

This perspective that has developed under the MP is highly speculative. There is still a considerable way to go in understanding how computation in human language is both efficient and optimal, how the governing principles for language constitute principles of efficient computation, how such principles relate to as yet unknown principles of neural organization, and finally how these latter principles "may be even more deeply grounded in physical law." (Note that the idea that some properties of biological systems can be explained by physical law originates in the work of d'Arcy Thompson (1942)—see Freidin & Vergnaud 2001 and Freidin & Lasnik 2011 for further discussion of this idea as it relates to generative grammar). While it is plausible that answers to the first two parts can be achieved, answers to the last two parts would require a double unification of linguistics and neuroscience on the one hand (see Poeppel & Embick 2005 for a discussion of the problems this involves) and neuroscience and physics on the other.

The study of modern generative grammar over the past sixty plus years has yielded an understanding of the nature of human language, its universal character, and the ways in which languages can vary, that is far beyond what preceded during several thousand years. The MP moves the field toward a deeper understanding by raising more challenging questions while at the same time imposing more rigorous standards on what might count as an answer.

Note

I am indebted to Howard Lasnik for insightful comments on the previous draft of this chapter.

References

Baker, M. 2001. *Atoms of language,* New York, NY, Basic Books.

Boeckx, C. 2003. *Islands and chains: Resumption as stranding,* Amsterdam, John Benjamins.

Borer, H. 1984. *Parametric syntax,* Dordrecht, Foris.

Chomsky, N. 1957. *Syntactic structures,* The Hague, Mouton.

Chomsky, N. 1964. The logical basis of linguistic theory. *In:* Lunt, H. (ed.) *Proceedings of the Ninth International Congress of Linguistics*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the theory of syntax,* Cambridge, MA: MIT Press.

Chomsky, N. 1970. Remarks on nominalization. *Indiana University Linguistic Club* Publications, reprinted in: Jacobs, R. A. & Rosenbaum, P. S. (eds.) *Readings in English transformational grammar*. Waltham, MA: Ginn and Co.

Chomsky, N. 1972. *Studies on semantics in generative grammar,* The Hague: Mouton.

Chomsky, N. 1973. Conditions on transformations. *In:* Anderson, S. & Kiparsky, P. (eds.) *A festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.

Chomsky, N. 1975. *The logical structure of linguistic theory*, New York: Plenum.
Chomsky, N. 1976. Conditions on rules of grammar. *Linguistic Analysis*, 2, 303–351.
Chomsky, N. 1977. On *wh*-movement. *In:* Culicover, P., Wasow, T. & Akmajian, A. (eds.) *Formal syntax*. New York, NY: Academic Press.
Chomsky, N. 1979. *The morphophonemics of Modern Hebrew*, New York: Garland.
Chomsky, N. 1980. On binding. *Linguistic Inquiry*, 11, 1–46.
Chomsky, N. 1981a. *Lectures on government and binding*, Dordrecht: Foris.
Chomsky, N. 1981b. Principles and parameters in syntactic theory. *In:* Lightfoot & N. Hornstein (ed.) *Explanation in linguistics*. London: Routledge.
Chomsky, N. 1986a. *Barriers*, Cambridge, MA: MIT Press.
Chomsky, N. 1986b. *Knowledge of language*, New York, Praeger.
Chomsky, N. 1988. *Language and problems of knowledge*, Cambridge, MA: MIT Press.
Chomsky, N. 1991. Some notes on economy of derivation and representation. *In:* Freidin, R. (ed.) *Principles and parameters in comparative grammar*. Cambridge, MA: MIT Press.
Chomsky, N. 1993. A minimalist program for linguistic theory. *In:* Hale, K. & Keyser, S. J. (eds.) *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.
Chomsky, N. 1994. Naturalism and dualism in the study of language and mind. *International Journal of Philosophical Studies*, 2, 181–200.
Chomsky, N. 1995a. Bare phrase structure. *In:* Campos, H. & P. Kempchinsky (ed.) *Evolution and revolution in linguistic theory*. Washington, D.C.: Georgetown University Press.
Chomsky, N. 1995b. *The minimalist program*, Cambridge, MA: MIT Press.
Chomsky, N. 1995c. Language and nature. *Mind*, 104, 1–61.
Chomsky, N. 2000. *New horizons in the study of language and mind*, NewYork: Cambridge University Press.
Chomsky, N. 2001. Derivation by phase. *In:* Kenstowicz, M. (ed.) *Ken Hale: A life in language*. Cambridge, MA: MIT Press.
Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry*, 36, 1–22.
Chomsky, N. 2007. Approach UG from below. *In:* Saurerland, U. & H.-M. Gärtner, (ed.) *Interfaces + Recursion = Language? Chomsky's minimalism and the view from syntax-semantics*. Berlin: Mouton de Gruyter.
Chomsky, N. 2008. On phases. *In:* Freidin, R., Otero C. & Zubizarreta M-L. (eds.) *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press.
Chomsky, N. & H. Lasnik. 1977. Filters and control. *Linguistic Inquiry*, 8, 425–504.
Fiengo, R. 1974. Semantic conditions on surface structure. Doctoral dissertation, MIT.
Fitch, T., M. Hauser, & N. Chomsky 2005. The evolution of the language faculty: Clarifications and implications. *Cognition*, 97, 179–210.
Freidin, R. 1978. Cyclicity and the theory of grammar. *Linguistic Inquiry*, 9, 519–549.
Freidin, R. 1992. *Foundations of generative syntax*, Cambridge, MA, MIT Press.
Freidin, R. 1995. Conceptual shifts in the science of grammar: 1951–1992. *In:* Otero, C. (ed.) *Noam Chomsky: critical assessments*. vol. 1, London: Routledge.
Freidin, R. 1997. Binding theory on minimalist assumptions. *In:* Rooryck, P. P. A. J. (ed.) *Atomism and binding*. Dordrecht: Foris.
Freidin, R. 1999. Cyclicity and minimalism. *In:* Epstein, S. D. & Hornstein, N. (eds.) *Working minimalism*. Cambridge, MA: MIT Press.
Freidin, R. 2004. *Syntactic structures* Redux. *Syntax*, 7, 100–126.
Freidin, R. & H. Lasnik. 2011. Some roots of minimalism in generative grammar. *In:* Boeckx, C. (ed.) *Handbook of linguistic minimalism*. New York: Oxford.
Freidin, R. and J.-R. Vergnaud. 2001. Exquisite Connections: some remarks on the evolution of linguistic theory. *Lingua*, 111, 639–666.
Harris, Z. S. 1951. *Methods in structural linguistics*, Chicago: University of Chicago Press.
Hauser, M., N. Chomsky, & T. Fitch 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.
Jackendoff, R. 1972. *Semantic interpretation in generative grammar*, Cambridge, MA: MIT Press.
Jackendoff, R. & S. Pinker. 2005. The faculty of language: What's special about it? *Cognition*, 95, 201–236.
Jenkins, L. 2000. *Biolinguistics: Exploring the biology of language*, New York: Cambridge University Press.
Kayne, R. 1975. *French syntax: The transformational cycle*, Cambridge, MA: MIT Press.
Kayne, R. 2000. *Parameters and universals*, New York: Oxford University Press.
Kayne, R. 2008. Some preliminary comparative remarks on French and Italian definite articles. *In:* Freidin

R., Otero, C., & Zubizarreta, M.-L. (eds.) *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press.
Lakoff, G. 1971. On generative semantics. *In:* Steinberg, D. S. & Jakobovits, L. A. (eds.) *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge: Cambridge University Press.
Larson, R., Viviane Déprez & Hiroko Yamakido (eds.) 2010. *The evolution of human language: Biolinguistic perspectives*, New York: Cambridge University Press.
Lasnik, H. 1991. On the necessity of binding conditions. *In:* Freidin, R. (ed.) *Principles and parameters in comparative grammar*. Cambridge, MA: MIT Press.
Lasnik, H. 1995. Last resort. *In:* Haraguchi, S. & Funaki, M. (eds.) *Minimalism and linguistic theory*. Tokyo: Hituzi Syobo.
Lasnik, H. 2000. *Syntactic Structures revisited: Contemporary lectures on classic transformational theory*, Cambridge, MA: MIT Press.
Lasnik, H. 2004. Conceptions of the cycle. *In:* Corver, N. & Chen, L.-S. (eds.) *WH-Movement: Moving on*. Cambridge, MA: MIT Press.
Lasnik, H. 2008. On the development of Case theory: Triumphs and challenges. *In:* Freidin, R., Otero, C., & Zubizarreta, M.-L. (eds.) *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press.
Lasnik, H.and M. Saito 1984. On the nature of proper government. *Linguistic Inquiry*, 15, 235–289.
Lasnik, H. and M. Saito 1992. *Move a*, Cambridge, MA: MIT Press.
Matthews, G. H. 1965. *Hidatsa syntax*, London: Mouton.
May, R. 1977. The grammar of quantification. Doctoral dissertation, MIT.
May, R. 1985. *Logical form: Its structure and derivation*, Cambridge, MA: MIT Press.
Mccawley, J. D. 1968. The role of semantics in grammar. *In:* Bach, E. & Harms, R. T. (eds.) *Universals in linguistic theory*. New York: Holt, Rinehart and Winston.
Newmeyer, F. 1986. *Linguistic theory in America: The first quarter-century of transformational generative grammar*, New York: Academic Press.
Nunes, J. 2004. *Linearization of chains and sideward movement*, Cambridge, MA: MIT Press.
Perlmutter, D. 1971. *Deep and surface constraints in syntax*, New York: Holt, Rinehart and Winston.
Phillips, C. 2010. Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* (forthcoming).
Pinker, S. and. R. Jackendoff. 2005. The nature of the language faculty and its implications for the evolution of language. *Cognition*, 97, 211–225.
Poeppel, D. and D. Embick. 2005. Defining the relation between linguistics and neuroscience. *In:* Cutler, A. (ed.) *Twenty-first century psycholinguistics: Four cornerstones*. Mahwah, NJ: Lawrence Erlbaum Associates.
Pollock, J.-Y. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365–424.
Postal, P. 1964. Constituent structure: a study of contemporary models of syntactic description. *International Journal of American Linguistics*, 30.
Quicoli, A. C. 1976a. Conditions on quantifier movement in French. *Linguistic Inquiry*, 7, 583–607.
Quicoli, A. C. 1976b. Conditions on clitic movement in Portuguese. *Linguistic Analysis*, 2, 199–223.
Rizzi, L. 1980. Violations of the *wh*-island constraint and the subjacency condition. *Journal of Italian Linguistics*, 5, 157–195.
Ross, J. R. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT.
Stockwell, R. P., Schachter, P., & Partee, B. H. 1973. *The major syntactic structures of English*, New York: Holt, Rinehart and Winston.
Stowell, T. 1981. Origins of phrase structure. Doctoral dissertation, MIT.
Uriagereka, J. 1999. Multiple spell-out. *In:* Epstein, S. D. & Hornstein, N. (eds.) *Working minimalism*. Cambridge, MA: MIT Press.
Thompson, D. W. 1942. *On growth and form*. Cambridge: Cambridge University Press.
Vergnaud, J.-R. 1977/2008. Personal letter to Howard Lasnik and Noam Chomsky. *In:* Freidin, R., Otero, C., & Zubizarreta, M.-L. (ed.s) *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press.

INDEX