Keynes, Post-Keynesianism and Political Economy

Essays in honour of Geoff Harcourt
Volume three

Edited by
Claudio Sardoni and Peter Kriesler

Routledge Frontiers of Political Economy

KEYNES, POST-KEYNESIANISM AND POLITICAL ECONOMY

Geoff Harcourt has had a major impact on the field of post-Keynesian economics, not only in his research but also in his teaching. Many of Harcourt's students have gone on to make valuable contributions in this field. This volume brings together contributions from thirty such former students, now established in academic institutions around the world. Their contributions focus on themes important to Harcourt's work, touching upon:

- history of political economy
- methodology
- economic theory
- applied analysis

This work is a valuable addition to the literature of political economy and a fitting tribute to a leading economist.

Claudio Sardoni is Associate Professor of Economics at the University of Rome 'La Sapienza'. He has published widely on Keynesian macroeconomic theory and on the relationship between the classical and Keynesian traditions. **Peter Kriesler** is an Associate Professor in Economics at the University of New South Wales, Australia. His publications include work on Kalecki and Keynes as well as an analysis of the Australian economy.

ROUTLEDGE FRONTIERS OF POLITICAL ECONOMY

1 EQUILIBRIUM VERSUS UNDERSTANDING
Towards the rehumanization of economics
within social theory
Mark Addleson

2 EVOLUTION, ORDER AND COMPLEXITY
Edited by Elias L.Khalil and Kenneth E.
Boulding

3 INTERACTIONS IN POLITICAL ECONOMY
Malvern after ten years
Edited by Steven Pressman

4 THE END OF ECONOMICS
Michael Perelman

5 PROBABILITY IN ECONOMICS
Omar F.Hamouda and Robin Rowley

6 CAPITAL CONTROVERSY, POST-
KEYNESIAN ECONOMICS AND THE
HISTORY OF ECONOMICS
Essays in honour of Geoff Harcourt, volume
one
Edited by Philip Arestis, Gabriel Palma and
Malcolm Sawyer

7 MARKETS, UNEMPLOYMENT AND ECONOMIC POLICY
andEssays in honour of Geoff Harcourt, volume
two
Edited by Philip Arestis, Gabriel Palma and
Malcolm Sawyer

8 SOCIAL ECONOMY
The logic of capitalist development
Clark Everling

9 NEW KEYNESIAN ECONOMICS/POST-KEYNESIAN ALTERNATIVES
Edited by Roy J.Rotheim

10 THE REPRESENTATIVE AGENT IN MACROECONOMICS
James E.Hartley

11 BORDERLANDS OF ECONOMICS
Essays in honour of Daniel R.Fusfeld
Edited by Nahid Aslanbeigui and Young Back
Choi

12 VALUE, DISTRIBUTION AND CAPITAL
Essays in honour of Pierangelo Garegnani
Edited by Gary Mongiovi and Fabio Petri

13 THE ECONOMICS OF SCIENCE
Methodology and epistemology as if economics really mattered
James R.Wible

14 COMPETITIVENESS, LOCALISED
LEARNING AND REGIONAL DEVELOPMENT
Specialization and prosperity in small open
economies

Peter Maskell, Heikki Eskelinen, Ingjaldur Hannibalsson, Anders Malmberg and Eirik Vatne

15 LABOUR MARKET THEORY
A constructive reassessment
Ben J.Fine

16 WOMEN AND EUROPEAN EMPLOYMENT
Jill Rubery, Mark Smith, Colette Fagan,
Damian Grimshaw

17 EXPLORATIONS IN ECONOMIC METHODOLOGY
From Lakatos to empirical philosophy of
science
Roger Backhouse

18 SUBJECTIVITY IN POLITICAL ECONOMY
Essays on wanting and choosing
David P.Levine

19 THE POLITICAL ECONOMY OF MIDDLE EAST PEACE
The impact of competing Arab and Israeli
trade
Edited by J.W.Wright Jr

20 THE ACTIVE CONSUMER
Novelty and surprise in consumer choice
Edited by Marina Bianchi

21 SUBJECTIVISM AND ECONOMIC ANALYSIS
Essays in memory of Ludwig Lachmann
Edited by Roger Koppl and Gary Mongiovi

22 KEYNES, POST-KEYNESIANISM AND POLITICAL ECONOMY
Essays in honour of Geoff Harcourt, volume three
Edited by Claudio Sardoni and Peter Kriesler

KEYNES, POST-KEYNESIANISM AND POLITICAL ECONOMY

Essays in Honour of Geoff Harcourt

Volume Three

Edited by
Claudio Sardoni and Peter Kriesler

London and New York

First published 1999
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

This edition published in the Taylor & Francis e-Library, 2005.

Routledge is an imprint of the Taylor & Francis Group

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Themes in post-Keynesian economics/[edited by] Claudio Sardoni and Peter Kriesler.
(Essays in honour of Geoff Harcourt: v. 3)
Includes bibliographical references and index.
1. Keynesian economics. 2. Economics—Methodology. 3. Economics—History. 4. Economic policy. 5. Harcourt, Geoffrey Colin. I. Sardoni, Claudio II. Harcourt, Geoffrey Colin. III. Series.
HB99.7.T44 1998
330. 15 – dc21 98 – 25293

ISBN 0-203-43827-2 Master e-book ISBN

ISBN 0-203-74651-1 (Adobe eReader Format)
ISBN 0-415-14042-0 (Print Edition)

CONTENTS

ILLUSTRATIONS

FIGURES

TABLES

CONTRIBUTORS

Jorge Thompson Araujo is an economist at the Economic Development Institute of the World Bank. He holds a Ph.D. in economics from Cambridge University and previously taught at the University of Brasilia, Brazil. He published several articles on economic growth, capital accumulation and functional distribution of income in journals such as *The Manchester School, Metroeconomica, Journal of the History of Economic Thought, Socio-Economic Planning Sciences* and *Journal of Post Keynesian Economics*.

Michelle Baddeley is Director of Studies in Economics at Gonville & Caius College, Cambridge. Her Ph.D. thesis analysed Keynesian and neoclassical ideas about rationality and expectations in investment decision-making. More recently, she has also been investigating European regional labour markets and this research has been published in journals such as *Applied Economics*.

Pulapre Balakrishnan is Country Economist for the Ukraine, Europe and Central Asia Regional Office, the World Bank, Washington DC. His main interest is in macroeconomics and he is the author, among other publications, of *Pricing and inflation in India* (1991).

Diana Barrowclough is an economics fellow at St John's College, Cambridge. Her research interests include the broad arena of public economics, particularly the recent introduction of quasi-markets and competition into the finance and provision of public goods. She is currently completing Ph.D. research at the University of Cambridge.

John F.Berdell is Assistant Professor of Economics at DePaul University, Chicago, Illinois. His research interests are in classical and international economics. He has published articles on the monetary theory and international political economies of Hume and Smith in the *Economic Journal, History of Political Economy* and *Review of Social Economy*. He is presently completing a volume on the classical approach to openness and growth.

Mauro Boianovsky currently teaches in the Department of Economics at the Universidade de Brasilia. He took his doctoral degree at Cambridge. He has published several articles on Wicksell, Pareto and other topics in journals like

History of Political Economy, Journal of the History of Economic Thought and the *European Journal of the History of Economic Thought*.

Giuliana Campanelli received her doctoral degree at Cambridge. She is currently Associate Professor of Economics at the University of Bologna. Her main publications are on growth and structural change and on early mathematical formulation of classical economics.

Giuseppe Ciccarone, Associate Professor of Economic Policy, currently teaches at the University of Rome 'La Sapienza' and took his Ph.D. at Cambridge. He has published a volume on the theoretical foundations of economic policy and is the co-author of a book on growth and distribution in Italy. He has also published several articles on economic policy games, on the theory of trade unions and on the welfare state.

Flavio V.Comim is currently completing a Ph.D. thesis at St Edmund's College, Cambridge, on the topic of 'Common-sense economies'. He has published on the history of economic thought, economic methodology and development economics.

Gianluca Cubadda is Associate Professor of Economic Statistics at the University of Campobasso, where he received his doctorate. He previously worked in the economic department of the Italian Central Statistical Office. His research interests include time series analysis and applied macroeconomics. He has published in international journals including *The Journal of Time Series Analysis* and *The Journal of the Italian Statistical Society*.

David Dequech is a Ph.D. student in economics at the University of Cambridge. His recent publications include: 'Different views on uncertainty and some policy implications', in P.Davidson and J.Kregel (eds) (1997) *Improving the Global Economy,* 'Uncertainty in a strong sense: meaning and sources', *Economic Issues* (1997); and 'A brief note on Keynes, unknown probabilities and uncertainty', *History of Economic Ideas* (1997).

Sergio Destefanis is Associate Professor of Economics at the University of Salerno, Italy. He took his doctorate at Cambridge. He has published several works on wage and price determination, dualism in the Italian economy and the economic analysis of efficiency. He has been awarded the Tarantelli prize for the best paper given at the 1995 meeting of the Italian Association of Labour Economics.

Stefano Fachin received his doctorate from Cambridge University and is Associate Professor of Economic Statistics at the University of Rome 'La Sapienza'. His research interests include econometric theory (analysis of structural breaks, applications of the bootstrap) and applied macroeconomics (economics of localisation, models of migrations) and has published in international journals including *The Oxford Bulletin of Economics and Statistics, The Journal of Forecasting,* and *Labour*.

Mario Garcia is a lecturer at the Universidad Nacional de Colombia and is currently completing his Ph.D. thesis at Wolfson College, Cambridge. He has published several articles on monetary theory and the history of Colombian theatre during the nineteenth century.

Jayati Ghosh is Associate Professor at the Centre for Economic Studies and Planning, Jawaharlal Nehru University, New Delhi. She received her doctorate from the University of Cambridge. Her research interests include problems of development and the international economy.

Roy Green is the Director of the Employment Studies Centre at the University of Newcastle, Australia. He graduated in law and economics from the University of Adelaide and took his Ph.D. at the University of Cambridge, where he was also a Research Fellow of Clare Hall. He is the author of *Classical Theories of Money, Output and Inflation: A Study in Historical Economics* (1992) and has written widely on industry and labour market issues.

Pierre Etienne Grégoire is currently working as an economist at the Ministère de l'Industrie du Commerce, de la Science et de la Technologic, Quebec, Canada. He studied at Jesus College, Cambridge, under the supervision of Geoff Harcourt from 1984 to 1987 and received his Ph.D. in 1989.

Prue Kerr has a Ph.D. from the University of Cambridge, and presently teaches at the University of New South Wales.

Richard Kozul-Wright, since completing his Ph.D. at Cambridge, has worked as an economist for the United Nations. He is currently at the United Nations Conference on Trade and Development, Geneva. He has published articles on international economics, economic development and the history of economic thought in the *Economic Journal, Cambridge Journal of Economics, Journal of Development Studies* and *Oxford Economic Papers*. He has recently co-edited with R.Rowthorn a collection of papers on *The Transnational Corporation and the Global Economy* (1997).

Peter Kriesler has a Ph.D. from the University of Cambridge which was supervised by Geoff Harcourt. He is currently an Associate Professor atthe University of New South Wales, and has published extensively in the history of economic thought, in post-Keynesian theory and on the economies of Australia, Japan and East Asia.

Rita McWilliams Tullberg, University of Cambridge, editor of *Alfred Marshall in Retrospect* (1990) and co-editor of *Alfred Marshall's Lectures to Women* (1995), is the author of numerous papers, book contributions, etc. on Alfred Marshall and on the history of women's higher education. She currently lives in Stockholm.

Robert Neugeboren is Lecturer in Economics and Allston Burr Senior Tutor in Cabot House (Assistant Dean), Harvard University. He completed his postgraduate studies at the University of Cambridge in England.

Francesco Nucci, currently lecturer in the Faculty of Statistics of the University of Rome 'La Sapienza', is a Ph.D. candidate at the Department of Economics of Columbia University. He previously worked as an economist in the research department of the Bank of Italy. His fields of interest are empirical macroeconomics and international economics.

Allen Oakley is Associate Professor of Economics at the University of Newcastle, Australia, where he has been since 1977. He is a graduate of the University of Adelaide, where he acquired his continuing research interest in the history and methodology of economics. His publications include a number of books on various themes in these fields, the most recent titles being *The Foundations of Austrian Economics from Menger to Mises* (1997); *Classical Economic Man* (1994); and *Schumpeter's Theory of Capitalist Motion* (1990).

Rod O'Donnell is Professor of Economics at Macquarie University, Sydney. A graduate of both economics and philosophy, he has written *Keynes: Philosophy, Economics and Politics,* edited *Keynes as a Philosopher-Economist,* and contributed articles to journals in both disciplines. He is currently working on a multi-volume edition of Keynes's remaining unpublished writings.

Terry O'Shaughnessy has a Tutorial Fellowship at St Anne's College, Oxford. He studied Engineering and Mathematics in Adelaide and took his Ph.D. in Cambridge. His research interests are in macroeconomics and applied welfare economics. Among his publications: 'Keynes's unemployment equilibrium: some insights from Joan Robinson, Piero Sraffa and Richard Kahn', (with G.C.Harcourt, in G.C.Harcourt (ed.) (1985) *Keynes and His Contemporaries;* ' The assessment: the balance of payments and international economic integration' (with Christopher Allsopp and Tim Jenkinson), *Oxford Review of Economic Policy* (1990);and 'Kahn on the economics of the short period', *Cambridge Journal of Economics* (1994).

Claudio Sardoni took his Ph.D. in Adelaide under the supervision of Geoff Harcourt and currently is Associate Professor of Economics at the University of Rome 'La Sapienza'. He has published *Marx and Keynes on Economic Recession* (1987) and articles on Marxian, Keynesian and post-Keynesian economics in various journals.

Keiran Sharpe is Lecturer in Economics at Queen's College, Oxford. He has previously studied and taught at the University of Sydney and at Cambridge University.

Pervez Tahir is Chief of the Economic Affairs Division, Government of Pakistan, Islamabad.

Elizabeth Webster is a research fellow at the Melbourne Institute of Applied Economic and Social Research, University of Melbourne. Her research interests include the interaction between the labour market and macroeconomic relationships, especially the effects of intangible investment activity.

Michael V.White is a senior lecturer at Monash University, and has written extensively in the history of economic thought.

FOREWORD

K.G.Arrow
Professor Emeritus, Stanford University

It is a pleasure to say a few words to introduce this book of essays in honor of Geoff Harcourt. As I recall, I first met this pleasant, yet firm young Australian scholar in the academic year 1963 – 4, when I was an Overseas Research Fellow at Churchill College and was allowed an office in the Faculty of Economics and Political Science (now renamed E.A.G. Robinson Hall). The then-existing intellectual conflicts among economists are already dimming to one's memory and hardly known to the current generation, but they were fierce and divisive. The issues took many different forms. A particularly virulent controversy was over the arcane question of 'reswitching'. In a world of many commodities, the concept of 'capital' was not clearly defined. If one defined it in some reasonable way, was it true that an increase in the rate of interest (or a decrease in the propensity to save) necessarily diminished the amount of capital in the system? Joan Robinson maintained stoutly that there was no necessary relation; she got analytical support from David Champernowne. Paul Samuelson and Bob Solow had models of the aggregate economy in which aggregate capital entered as a variable; naturally, with the usual concavity conditions, the demand for capital was a decreasing function of the rate of interest. Samuelson even got a graduate student to 'demonstrate' that this result would hold in a disaggregate model; but Luigi Pasinetti showed this was wrong.

This was the state of the 'two Cambridges' controversy in 1963 – 4, when Bob Solow and I were both visiting Cambridge. Tempers ran high; Joan regarded us as lackeys of capitalism, and James Meade was not on speaking terms with Joan and others who felt they were carrying the Keynesian torch in an environment made more hostile by the revival of neoclassical economics.

But the Cambridge environment was made tolerable and even something constructive for the visitors by the sweet temper and constructive scholarship of Geoff Harcourt (and, in a very different way, by the intellectual breadth of the late Nicholas Kaldor). Geoff had a very broad-ranging set of interests in economics. I remember his interest in the history of economic thought and, in particular, in the methodology of that history, today as thencontroversial. He introduced and defended a young scholar at a seminar, one who was advocating a method of 'rational reconstruction', very far removed from the 'unmasking'

approach so common in Cambridge at that time (e.g. neoclassical economics in deliberate obfuscation designed to conceal the evils of the present system). He was concerned with the logic of all points of view and with their abilities to explain the broad facts of the present system. As a result, Geoff expressed much more clearly than anyone else the nature of the differences in viewpoint between the contending parties, emphasising their analytic content and the precise ways they agreed and disagreed, and playing down their expressions as rigid doctrines.

I do not mean that Geoff was a wishy-washy eclectic; he certainly had firm convictions and maintained them. I only came to realise later the extent to which his concerns were motivated by strong humanitarian and egalitarian values derived from his religion. Regardless of their source, these are values which both sides in the 1960s controversies hold high, and which have been under steady attack in the last twenty years. Geoff's firm convictions have been a pillar of fire in the night.

INTRODUCTION

C.Sardoni and P.Kriesler

This volume is the third in honour of Geoff Harcourt's 65th birthday. The original project to honour Geoff was planned as a single volume, but there were so many friends, colleagues, ex-students and students of his who expressed their wish to contribute to the project that the people who first organised the *Festschrift* soon realised that they needed two volumes. Even this proved not to be enough: too many people would have had to be left out if the two volumes were to be a publishable size. Thus P.Arestis, M.Sawyer and G.Palma proceeded to edit the first two volumes (1997), while C. Sardoni and P.Kriesler, who were both Ph.D. students of Geoff's, set out to edit the present third volume. What distinguishes this volume from the others is that all the contributors have been, at some time or another, Geoff's students.

This volume contains twenty-eight essays. The large number of contributions is a good indicator of Geoff Harcourt's popularity among those who studied, or are studying, under his supervision as graduate or postgraduate students. There are not many university teachers who in the course of their career have attracted so much admiration, respect and affection as Geoff has, nor have inspired as many students to take up the challenge of an academic career.

Another characteristic of this volume is the enormous geographical diversity of the contributors, literally from all over the world: Australia, New Zealand, Canada, USA, India, Pakistan, Brazil, Italy, Britain and Sweden. Although Geoff Harcourt has taught primarily in Australia, Canada and Britain, he has attracted students from a wide range of locations. This has been due not only to the reputation of his academic works, but also to the impression of friendliness, warmth and intellectual availability that he always left behind him in the many places that he visited to give seminars or lectures.

The range of topics considered in the essays in this volume is also revealing of the significant impact that Geoff Harcourt has had on younger economists. They range from history of political economy to applied analysis, macroeconomics and other theoretical issues. The editors have dividedand collected the essays under two subheadings of the general theme of political economy. The first, 'History of political economy and methodology' covers diverse issues in political economy.

The papers are ordered in chronological order of their subjects. The dominant themes of these essays represent the main concerns of Geoff's academic life, with chapters on the history of thought, particularly on Joan Robinson, Keynes and Kalecki; as well as contributions on post-Keynesian themes. The second part of this volume reflects Geoff's great concern with practical and applied matters. He has written extensively on policy issues, and been an advisor to the Australian Labor Government.

Regardless of the specific topics that authors have chosen to deal with, the essays all share a basic characteristic. The authors treat their topics with the evident intention of providing a contribution to the development of a more satisfactory approach to the understanding and improving of the economy. This is very clear and obvious for all of the essays. The authors have not embarked on the reconstruction of debates of the past, the discussion of methodological issues, or the analysis of theoretical or applied problems, for the sake of erudition. From different perspectives, they are all trying to contribute to developing economics along lines that allow a more satisfactory and comprehensive understanding of the world in which we live, in order to better it. Geoff Harcourt's influence on his students choosing this perspective is evident. Sometimes at different stages and sometimes simultaneously, Geoff has delved into history, methodology, theory and applied analysis, always with the objective of helping to construct a better and more just way to organise the production and distribution of wealth. In his view, such an objective cannot be realised without this comprehensive approach to economics.

Geoff Harcourt has tried to accomplish the task that he set for himself by taking inspiration from his 'heroes', the great names of the Cambridge tradition: Keynes, Sraffa, Joan Robinson, Kaldor and Kalecki. He is one of the most prominent representatives of post-Keynesian economics. Many of the authors in this volume, not surprisingly, share with Geoff the same post-Keynesian perspective; but this is not true for everybody. Some authors look at their topics from different theoretical perspectives. What is important is that, regardless of their preferred theoretical stance, each author retains this same basic attitude toward economics. We believe that this is a sign of a teacher's success.

Serious teaching is not about transmitting a set of unquestionable truths or recruiting a large number of faithful followers of one's beliefs and values. Serious teaching is about helping people make their own choices by retaining open-mindedness, intellectual curiosity and integrity. From this point of view, Geoff has offered a lot of himself to all those with the good fortune to have had him as a teacher and friend. Two things that, for Geoff, almost always make a joint product.[1]

NOTE

1 A more detailed discussion of Geoff's life and contribution is contained in the introduction to the first two volumes.

REFERENCE

Arestis, P., Palma, G. and Sawyer, M. (eds) (1997) *Capital Controversy, Post-Keynesian Economics and the History of Economics: Essays in Honour of Geoff Harcourt,* vol. 1 and *Markets, Unemployment and Economic Policy: Essays in Honour of Geoff Harcourt,* vol. 2, London and New York: Routledge.

Part I

HISTORY OF POLITICAL ECONOMY AND METHODOLOGY

1
MANUFACTURING THE SMITHIAN PARADOX OF VALUE

Michael V.White

INTRODUCTION

It has been suggested that paradoxes, which economists regard as puzzling outcomes, can be understood as 'a normal aspect of ongoing inquiry' in that they act as a 'stimulus to further research'. An illustration of this process is provided by the water and diamonds paradox,[1] which Adam Smith presented in the *Wealth of Nations* as a rhetorical paradox, a device where 'terminological fuzziness…[is used] to excite puzzlement and interest in the reader' (De Marchi 1987:796 – 7). This characterisation of the significance of Smith's paragraph is consistent with the approach taken by those historians of economics who depict the paradox as providing 'the starting-point for the theorising of economists of the later nineteenth century which finally led to the marginal utility doctrine' (Roll 1966: 156).

A synthesis of accounts which argue that the paradox was a, if not the, crucial explanatory factor in explaining the advent of the (British) 'marginalist revolution', might read as follows. Until 1871, the explanation for the 'discrepancy' between value in use and value in exchange remained a mystery because the classical economists were 'incapable' of resolving it (Ekelund and Hebert 1990:107). Smith was actually 'puzzled' by his own statement[2] and, 'hampered' by a failure to use the calculus, he and the Classical school were unable to grasp that the correct solution required a distinction between the marginal and total valuation of a commodity (*ibid.:* 161, 325; also Galbraith 1989:65). Without that solution, Smith effectively set aside any discussion of value in use and hurried on to 'explain only' the determination of exchange value (Roll 1966:156; also Schumpeter 1963: 300; Galbraith 1989:66; Ekelund and Hebert 1990:107). Unable to understand the 'existence and significance' of marginal valuations (Ekelund and Hebert 1990:107), other pre-marginalist economists concentrated on 'costof production as the explanation of price' (Cooter and Rappoport 1984: 510). Although a basis for the solution was suggested early in the nineteenth century by Jeremy Bentham (Hutchison 1956)

and Nassau Senior had the 'key' to the solution 'in his grasp' (Ekelund and Hebert 1990:161), it was generally 'puzzled over for nearly a hundred years in the economics literature' (Schmidt 1992:1). The denouement came in 1871 when William Stanley Jevons finally 'unlocked' the paradox with the distinction between marginal and total utility in his *Theory of Political Economy* (Ekelund and Hebert 1990:358; also Cooter and Rapport 1984:510).[3]

This synthesis would not, of course, command universal assent by historians of economics,[4] even by those who accept that a paradox existed before the publication of Jevons' *Theory*. Indeed, it has been argued recently that, because Jevons did not discuss the issue until the second edition of his *Theory* in 1879, the paradox 'figured only indirectly in the conceptual genesis' of the marginal utility theory. Nevertheless, the treatment of water and diamonds in pre-marginalist texts is still depicted as a 'longstanding anomaly' or 'paradox', a 'formal solution' for which was only provided by Jevons (Schabas 1990:23 – 4, 146). In common with the histories it effectively criticises, this account assumes that there was a general uniform or stable reading of the *Wealth of Nations* paragraph as a paradox which could not be explained.

The principal purpose of this chapter is to show that there was no such stability before 1871. The lack of stability can be demonstrated in two ways. First, as explained in the next section, those economists who accepted Smith's distinction between value in use (VIU) and value in exchange (VIE) did not regard it as a paradox. Nor did their use of the 'Classical' definition of utility, which underpinned the distinction, prevent them from providing an explanation for the relative prices of commodities such as water and diamonds. It was possible, moreover, to present the distinction between VIU and VIE as compatible with two different general explanations of prices—one based on 'cost of production' and the other based on 'scarcity' (the latter argument is often regarded as a forerunner of marginalist theory). The second reason for the lack of stability, considered in the third section, is that there was significant dissent from the terms of Smith's discussion, especially regarding the definition of utility. Again, however, there was no mention of a paradox. Instead, the most influential variant of the critique argued that Smith's discussion was incoherent.

The appearance of the paradox is explained in the final section, which shows that, in the first (not the second) edition of his *Theory,* Jevons argued that Smith's distinction between VIU and VIE could be read in terms of the difference between total and marginal utility. While this account was consistent with the redefinition of utility by Smith's critics, it was also quite different. For it depended upon a theory of human action which was only produced after 1850. Moreover, rather than reading Smith as incoherent,Jevons then presented him, for the first time, as outlining a 'paradox'. The stabilisation of Smith's discussion as the classical paradox of value was thus an artefact of British marginalism.

WHAT PARADOX?

In texts of political economy, a distinction between the usefulness or value of commodities (where they were graded in terms of the 'real' or 'intrinsic wants of mankind') and their value in exchange as price, was made by Barbon and Locke in the late seventeenth century. While their discussion could draw on a much older literature concerning use and price,[5] they were followed by Mandeville and Jocelyn early in the next century, the latter using water and diamonds as examples. The same example to illustrate a distinction between VIU and VIE can be found in *Money and Trade Considered* (1705) by John Law and *An Essay on Money and Coinage* (1757) by Joseph Harris.[6] Adam Smith had both texts in his library and used that example in his *Lectures on Jurisprudence* delivered in the 1760s, where he argued that the difference in the prices of water and diamonds could be explained by their relative abundance or scarcity (Smith 1978:333 – 4, 487).

In the *Wealth of Nations* (WN), having considered how 'money has become in all civilised nations the universal instrument of commerce', Smith posed the question of 'what rules determine what may be called the relative or exchangeable value of goods' (Smith 1976:44). He then observed, in the passage which was subsequently read as the classic statement of the diamonds and water paradox, that

> The word VALUE…has two different meanings, and sometimes expresses the utility of some particular object, and sometimes the power of purchasing other goods which the possession of that object conveys. The one may be called 'value in use'; the other, 'value in exchange'. The things which have the greatest value in use have frequently little or no value in exchange; and, on the contrary, those which have the greatest value in exchange have frequently little or no value in use. Nothing is more useful than water; but it will scarce purchase anything; scarce anything may be had in exchange for it. A diamond, on the contrary, has scarce any value in use; but a very great quantity of other goods may frequently be had in exchange for it.
>
> (*ibid.:* 44 – 5)

Smith then indicated that the analysis which followed was concerned with identifying 'the principles which regulate the exchangeable value of commodities' (*ibid.:* 46). His reference to VIU and VIE was thus 'solely tobring out' rather than to 'establish any distinction between the two', let alone to state a 'paradox' (Kaushil 1973:61n).[7] This point is reinforced by Smith's subsequent comment that the prices of precious metals, such as gold, were due to their utility, their beauty and their scarcity, while the demand for precious stones depended on their beauty: They are of no use, but as ornaments; and the merit of their beauty is greatly enhanced by their scarcity or by the difficulty and expense of getting

them up from the mine.' Indeed, the prices of precious stones and metals 'is regulated all over the world by their price at the most fertile mine in it' (*ibid.:* 189 – 91; also 563). This reference and other passages in WN indicate that Smith used the word 'utility' and hence defined VIU by referring to a hierarchical ranking of commodities in terms of their relative importance in meeting 'the wants of mankind'.[8] With the hierarchy characterised in terms of necessaries, conveniences and luxuries, it formed the basis for the Classical definition of utility which Malthus summarised as the 'quality of being serviceable or beneficial to mankind. The utility of an object has generally been considered as proportional to the necessity and real importance of these services and benefits'. Malthus did note that, with VIU considered as 'synonymous with utility', confusion was possible over the meaning of value so that the term VIU 'rarely occurs in political economy, and is never implied by the word value when used alone' (Malthus 1963:234 – 5; see also McCulloch 1864:3, 4). Nevertheless, the distinction between VIU and VIE was important because it could be used to compare 'that which is useful and that which is merely high-priced…that which is calculated to satisfy the acknowledged and general wants of mankind, and that which may only be calculated to satisfy the capricious wants of the few' (Malthus 1963:19).

Ricardo had used the same reading in his *Principles,* which began by registering agreement with Smith on the distinction between VIU and VIE (Ricardo 195la:11 – 12). The distinction is evident in his subsequent comment that the 'benefits of commerce' lay in the acquisition of 'not more valuable objects, but more useful ones' (*ibid.:* 264n) and was particularly important in Chapter 20 where J.B.Say (see below) was criticised for effectively conflating the analysis of 'riches' or wealth with VIE by arguing that VIE 'is the measure' of a commodity's utility (*ibid.:* 282). For Ricardo, if a change in the technique of production meant that two sacks of corn, instead of one, could be produced with the same labour input, their value would be unchanged while there would be 'double the quantity of riches—double the quantity of utility—double the quantity of what Adam Smith called value in use, but not double the quantity of value' (*ibid.:* 281; see also 283). Malthus applauded Ricardo's critique of Say's explanation of riches and value despite his strong disagreement with Ricardo's use of a labour theory of value (evident in the quotation above) to explain the long period or natural prices of commodities.[9]

Ricardo's analysis of natural prices provides one example of the differentexplanations of commodity prices which were considered compatible with the distinction between VIU and VIE. For Ricardo, utility (VIU) was an 'absolutely essential' precondition, but could not be a determinant of VIE because the natural prices of commodities were determined either by their 'scarcity alone' or by 'the quantity of labour required to obtain them'. He restricted the commodity domain in the *Principles* to the latter group, on the ground that '[b]y far the greater part of those goods which are the objects of

desire, are procured by labour' (Ricardo 1951a:11 – 12). The domain restriction can explain why there was no discussion of the VIE of water in the *Principles*. Ricardo followed Smith in assuming that, 'under ordinary circumstances', water had no price (*ibid.:* 11, 287) and, elsewhere, posed the rhetorical question, 'Why is water without value, but because of its abundance?' (Ricardo 1951b:221). The absence of discussion of a price for water in the *Principles* can thus be explained by its categorisation as scarcity-determined. It is possible that Ricardo regarded diamond prices as determined in the same way. This is, of course, a different matter from claiming that he was puzzled by their relative prices.[10]

The difficulties which Ricardo acknowledged with a simple labour theory of value were brushed aside by J.R.McCulloch (cf. McCulloch 1864:274) and his explanation of long period prices differed significantly from Ricardo's.[11] Nevertheless, as late as the fifth edition of his *Principles,* McCulloch kept to the distinction between VIE and VIU, which he illustrated with water and gold prices. Water was 'indispensable to existence, and has, therefore, a high degree of utility, or of "value in use"', but a low VIE 'in most places' because of the small extent of labour or 'exertion' required to obtain it. Gold had a comparatively high VIE because, although 'of little utility', it existed 'in limited quantities' and required much labour and exertion to obtain it (*ibid.:* 3, 4). For McCulloch, it was possible to explain a positive price for water by reference to '*cost of production* [which] is the grand regulator of price' (*ibid.:* 252 OE).[12] William Whewell then produced an account which seems generally consistent with McCulloch's, in that he argued that market prices were explained by supply and demand, while the long period prices of most commodities were governed by their cost of production (Whewell 1971:8 – 10; 1862:44 – 52). After quoting Smith on water and diamonds in his *Lectures on Political Economy* delivered at Cambridge in 1861, Whewell argued that, given an article was useful or provided 'enjoyments', its VIE depended on whether its 'procurement' required 'the labour of man' to produce it:

> Utility may therefore be considered as the sole cause of *value in use,* whilst *value in exchange* may be produced by any circumstance which renders the possession of an object so difficult of attainment, and at the same time so desirable, that men are willing to give something in exchange for it. Thus not only utility but beauty, curiosity,fashion, rarity, and many other qualities may create exchangeable value; and it is to this value that, in political economy, we chiefly confine our attention.
>
> (Whewell 1862:47 – 8 OE)[13]

Other explanations for price formation were considered compatible with Smith's distinction between VIU and VIE. In his 1846 and 1847 Oxford lectures, for example, Travers Twiss referred to the distinction and then described the 'principles…[which] regulate the exchangeable value of commodities' with what

appears to be a simplistic adding-up theory (Twiss 1847:166 – 7). More striking was the way that Richard Whately's *Easy Lessons* rejected McCulloch's argument that cost of production was the grand regulator of price and yet still accepted the distinction between VIU and VIE. Whately noted that it was necessary to distinguish between whether a commodity was 'useful' and whether it had *'value'* in exchange (Whately 1853:27 OE). VIE required that an article be an object of *'desire'* and that it be exchangeable, so that its owner 'can *part with*' it. The principal determinant of VIE was, however, the *'scarcity'* of a commodity, that is, whether it was '*limited* in supply, [so] that it is not to be had for nothing'. While a pearl or a diamond was of 'great value' because it was 'scarce…very *beautiful*' and constituted 'a *sign of wealth*', it was 'of no *use* but to make an ornament for a person' (*ibid.:* 27, 28, 29, 30 OE). However, while the scarcity of a commodity could be equated to a large extent with the labour cost of its acquisition, labour did not create value. It was the high price which 'causes' people to labour:

> It is not…labour that makes things valuable, but their being valuable that makes them worth labouring for. And God, having judged in his wisdom that it is not good for man to be idle, has so appointed things by his Providence, that few of the things that are desirable can be obtained without labour. It is ordained for man to eat bread in the sweat of his face; and almost all the necessaries, comforts, and luxuries of life, are obtained by labour.
>
> (*ibid.:* 32 – 3)

This argument, which can also be found in Whately's *Lectures,*[14] possibly owed a good deal to Nassau Senior (Bowley 1937:106 – 7), who used it in his *Outline* with the illustration of a pearl picked up from the 'sea-shore' (Senior 1965:24). It will be noted, however, that while Senior followed Say in arguing that VIU and VIE were closely connected (see below), Whately restated what Malthus (1963:20) was to describe as the universal custom of distinguishing between VIU and VIE.

The Classical definition of utility was used to designate a hierarchy of commodities in terms of the comparative usefulness of the different 'wantsof mankind', so that the utility of a particular commodity might be only one part of the explanation for those wants and desires. With the exception of Senior, the authors considered in this section were thus able to refer to the distinction between VIU and VIE, to illustrate it with water and diamonds (or pearls or gold) and to explain those prices. This was despite the marked differences in their explanations of VIE. Any suggestion that Smith's discussion of water and diamonds was understood as a 'paradox' which could not be explained before 1871 is simply a myth.[15]

‘INCONCEIVABLE AND MONSTROUS’: THE CRITIQUE (S)

It was noted above that Ricardo criticised Say’s argument that, with utility defined as the ‘inherent fitness or capability of certain things to satisfy the various wants of mankind’, to create utility was to create wealth and ‘value constitutes wealth’. Since price was the measure of value, value was the measure of utility (Say 1964:62). This argument was reworked by other critics of the classical distinction between VIU and VIE, using three steps which can be found in Senior’s *Outline*.

The first was to redefine utility so that it became a portmanteau term encompassing all wants and desires which in turn explained the demands for various commodities. Utility could now only mean the ‘ “desiredness”, or the capacity of a good or service to satisfy a want, of whatever kind’, a definition which was ‘clearly distinct’ from the Classical designation of ‘usefulness or fitness for a purpose’ (Black 1967:776; see also Bharadwaj 1978:256). Senior defined utility as ‘the power, direct or indirect, of producing pleasure, including under that term gratification of every kind, or of preventing pain, including under that term every species of discomfort’. While utility was ‘generally used to express the quality of preventing pain or of indirectly producing pleasure, as a means’, Senior ‘extended’ the usual (English) understanding of the term to include ‘all those things which produce pleasure directly’ (Senior 1965:6).[16] Given this, the second step was to link the new definition of utility directly with the determination of VIE. Here, Senior argued that a commodity’s VIE depended principally on its transferability (and hence exchangeability), the limitation in its supply (the ‘comparative force of the obstacles which limit the respective supplies of the objects compared’), and on demand which was explained by utility (Senior 1965:8, 11).

The third step in the critique was to argue that, with the new definition of utility, the classical reference to VIU could not explain VIE. For Say, ‘what is good for nothing…[men] set no price upon’ (Say 1964:62) while for Senior, ‘no man would give anything possessing the slightest utility for a thing possessing more, and even an exchange of two useless things would be,on the part of each party to the exchange, an act without a motive’ (Senior 1965:6 – 7). Say added the further claim that the Classical definition of utility was unacceptable because it introduced moralistic elements into the discourse: ‘whether or no the value mankind attach to a thing be always proportionate to its actual utility’ depended on the analyst’s ‘comparative judgement, intelligence, habits, and prejudices’. Such evaluations, however, were essentially those of a moralist, whereas political economy assumed the ‘appreciation of benefits’ to be ‘one of the *data* of its reasonings’ (Say 1964: 62 OE).[17]

Following Senior less than a decade later, Thomas Banfield felt able to dismiss the classical distinction in a few lines. Since the ‘origin of value’ depended on the

> supply of and demand for other objects, [it] does not admit of any distinction being established between value in use and value in exchange. When freedom of exertion and security of property exist in the full sense of the terms, value in use and value in exchange coincide.
>
> (Banfield 1848:17 OE)

The argument was also taken up by Thomas De Quincey and John Stuart Mill in what was to become the most influential version of the critique before 1871. Although this version relied on the three basic components of the critique discussed above, it also differed in two significant ways. The first was that references to supply and demand, underpinned by references to utility and scarcity, could not provide a general explanation for the VIE of all commodities.[18]. The second was an explicit criticism of, and focus on, Smith's discussion of water and diamonds, which recast the third point of the Say/Senior critique described above. The point of departure here, however, had been mapped in 1801 by Jeremy Bentham with his *True Alarm*.

Bentham claimed (Stark 1954:86 – 8) that Smith had divided 'all articles' into those which had VIU but no VIE, and those which had VIE but no VIU. The examples of water and diamonds were, however, 'ill chosen'. Water, for example, could have VIE,[19] although it was possible to explain why that might not be the case: 'If the whole quantity required is available, the surplus has no kind of value'. But the real problem was with diamonds. While they were 'fancy' values (and therefore 'neither essential nor invariable' unlike water), the 'fair sex' could have told Smith that it was quite wrong to argue that they had no VIU.[20] Moreover, to suggest that an article could have VIE but no VIU was absurd: 'who would give away a useful thing for another that could never have utility for anybody?' (*ibid.:* 87, 88).

If this critique depended on the argument that VIU could be used only to designate personal valuation of commodities in exchange, it also caricatured the relevant paragraph from WN since Smith had not divided 'all' commodi ties in the way that Bentham claimed. The caricature was then replicated by De Quincey in his *Logic of Political Economy* (1844) which illustrated the difference between VIU and VIE in a novel way by quoting the Roman author Plautus on the haggling between a brothel proprietor and customer.[21] The remainder of the argument, however, that '[i]n general…value in use as excluding value in exchange, has no place in political economy', was pure Bentham with some additional polemic. Smith's diamonds and water paragraph was a 'hyperbolic extravagance', the diamonds case 'a blank impossibility' ('inconceivable and monstrous') because it entailed that, in an exchange, a person 'will give more than the *maximum* which he *would* give' as indicated by the VIU. If diamonds had no use there could be no sale (De Quincey 1897:144 – 5, 187 – 8 OE). De Quincey added, following Say, that Smith's 'total misapprehension' of VIU depended on a 'puerile' moralism. Political economy

could not distinguish between 'the useful and the noxious' since use 'means the capacity of being…applied to a purpose…however monstrous, pernicious, or even destructive to the user' (*ibid.:* 120 – 1, 189 – 90 OE). Mill then indicated his agreement with De Quincey in his *Principles,* arguing that Smith had employed VIU in the 'sense in which use is opposed to pleasure. Political economy has nothing to do with the comparative estimation of different uses in the judgement of a philosopher or a moralist. The use of a thing, in political economy, means its capacity to satisfy a desire or serve a purpose' (Mill 1909: 437).

For Bentham, De Quincey and Mill, the discussion of water and diamonds in WN entailed a logical contradiction which led to absurdity. Two sources suggest that this critique was not without effect before 1871. In 1863, a reviewer of Whewell's *Lectures* noted that Whewell had cited Adam Smith on water and diamonds 'without any warning of the important error which it is now acknowledged to contain'.[22] Then, in 1870, Alfred Marshall opened his 'Essay on value' with a discussion of Smith on VIU and VIE, repeating Mill's argument that, by making VIU depend on utility, Smith 'makes himself the judge of what is useful to other people and introduces unnecessary confusion'. While Bentham, De Quincey and Mill focused on Smith's discussion of diamonds, Marshall argued that the analysis of water was also incorrect because, if a particular amount had no VIE, it could have no VIU. Moreover, if a consumer refused to pay for that amount of water, knowing that it could be obtained 'gratis' from elsewhere, this showed that Smith's 'error' when discussing VIU was that he had failed to analyse the importance for the determination of price of the terms on which alternative supplies of a commodity could be obtained. VIU should thus be defined as 'the amount of pleasure (a consumer) would derive from obtaining (a commodity) or the pain he would undergo from losing it' (Whitaker 1975: 125).

In considering the significance of Bentham's critique, it is thus misleading to argue that Smith's discussion of water and diamonds 'had asignificant influence on orthodox theorising in Britain for most of the next century until Jevons' (Hutchison 1978:14), without mentioning De Quincey and Mill. Their strand of the critique can in part be explained by the perceived need, *contra* Ricardo, to incorporate 'scarcity' commodities in a general theory of VIE. Criticism of Ricardo in that regard was evident in the work of Bailey (1967: Chapter XXI) and Senior (1965:24) to which both De Quincey and Mill responded (De Vivo 1981; Groenewegen 1982). Nevertheless, despite the break with Smith and Ricardo, a restatement, elaboration or reworking of Bentham's critique did not 'clearly point in the direction of the marginal utility theory' (Hutchison 1956: 291) as Edgeworth effectively noted when discussing De Quincey's analysis (Edgeworth 1987:813). Even more revealing in this regard is that, while Marshall explained VIU in terms of the pleasure and pain facing a consumer, and his 1870 manuscript contains some indications of the subsequent analysis of

consumer rent/surplus (Whitaker 1975:122, 130 – 1) he did not use a marginal utility theory to explain a downward sloping market demand curve (*ibid:* 122, 145). The use of that theory came only after he had read Jevons' *Theory of Political Economy* in 1871.[23] As will be shown below, Jevons' discussion of Smith owed nothing to the claim that the WN paragraph was incoherent. This suggests that, rather than being 'a step towards the later resolution of the [diamonds and water] paradox in terms of "total" and "marginal" utility' (Bharadwaj 1978:256), the De Quincey/Mill critique reached its terminus in Marshall's non-marginalist discussion of VIE in 1870.

THE JEVONIAN PARADOX

In the *Theory of Political Economy* (TPE), Jevons defined '*utility*…[as] the abstract quality whereby an object serves our purposes and becomes entitled to rank as a commodity. Whatever can produce pleasure or prevent pain *may* possess utility' (Jevons 1871:45 OE). This approach, where utility could only be specified in terms of possession, was clearly identified with that of Say and Senior (*ibid.:* 45, 52). Adam Smith's reference to VIU and VIE was, however, presented as broadly consistent with, or at least explicable within, the marginalist theory, although it required some criticism, clarification and emendation. In particular, like almost all political economists, Smith had 'failed to discriminate' between the final degree of (i.e. marginal) utility and total utility:

> From this confusion has arisen much perplexity. Many of those commodities which are the most useful to us are esteemed and desired the least. We cannot live a day without water, and yet in ordinary circumstances we set no value on it. Why is this? Simplybecause we usually have so much of it that its final degree of utility is reduced nearly to zero.
>
> (Jevons 1871:62)

Having used the word 'value' in this context, Jevons complained about the

> thoroughly ambiguous and unscientific character of the term…Adam Smith noted the extreme difference of meaning between *value in use* and *value in exchange;* and it is usual for the best writers on Economy to caution their readers against the confusion to which they are liable.
>
> (*ibid.:* 81 OE)

To avoid confusion, Jevons proposed to 'substitute the unequivocal expression—*ratio of exchange*', providing the following example of confusion which followed the use of 'value':

> there is a certain sense of esteem, or desirableness, which we may have with regard to a thing apart from any distinct consciousness of the ratio in which it may exchange for other things…[T]his distinct feeling of value is probably identical with the final degree of utility. While Adam Smith's often quoted *value in use* is the total utility of a commodity to us, the *value in exchange* is defined by the *terminal utility,* the remaining desire which we or others have for possessing more.
>
> (*ibid.:* 157 OE)

This point was crucial because, when depicted in the 'static' analysis of market prices in Chapter IV of TPE, the exchange ratio (VIE) 'depends solely on the final degree of utility', which in turn depended upon the quantity available of a commodity, i.e. its relative scarcity or abundance (*ibid:* 160). As he made clear to J.Shadwell in 1872, when explaining how market prices reflected the 'laws of variation of utility' in TPE, 'the main point of difference from Adam Smith was the distinguishing of the *degree of utility from the total amount of utility* (Black 1972 – 81:III, 254 OE).

It was noted above that Jevons identified his approach with Say and, especially, Senior in that, given exchangeability, the general explanation for VIE, understood as market price, depended on relative scarcity and utility/desire (Jevons 1871:156; 1878a:13 – 16). Indeed, he followed Say 'and several other writers, in regarding value as wholly dependent on utility' (Jevons 1875:16) This link was illustrated in the unfinished *Principles of Economics* where, echoing Senior (1965:11, 13), Jevons observed that the

> desire for a beautiful and distinguished ornament constitutes the utility, and this rests very much upon the extreme scarcity of fine large diamonds. Could they be manufactured in any desired quantity, people would soon cease to desire more as urgently as they do now.
>
> (Jevons 1905:62)

The link is evident also in his primer, *Political Economy,* where the discussion of the relation between scarcity, VIE and labour costs echoes that of Senior and Whately (Senior 1965:24; Whately 1853:26 – 33; Jevons 1878a: 101 – 3).

The argument in Senior's *Outline* was compatible with Jevons' analysis insofar as it described economic actions as driven by pleasure and pain and characterised market prices in terms of scarcity. However, Jevons' account also required a marginal utility theory so as to explain the distinction between total and marginal utility. This cannot be found in the *Outline* since Senior's explanation of utility and of behaviour was incompatible with the marginalist theory (White 1992). All the sufficient conditions can, however, be found in the *Natural Elements of Political Economy* by Richard Jennings which, when published in 1855, introduced a marked break in the depiction of human behaviour in political

economy by drawing on the new discourse of physiological psychology. In 1860, that text provided Jevons with the functionalist theory of behaviour of consumption and work/labour which he relabelled the theory of marginal utility and disutility (White 1994). Jennings' basic framework also suggested how water and diamond market prices could be depicted in terms of marginal and total utility.

Like Jevons, Jennings distinguished between the statics and dynamics of political economy. Dynamics considered fluctuations in output and employment, while statics was concerned, in part, with explaining market prices by the incremental 'value' which actors placed on successive units of a commodity according to its relative 'abundance'. (At one point, Jennings seemed to equate utility with value in consumption, which would have facilitated Jevons' relabelling of the analysis [Jennings 1855:246; Jevons 1871: 65 – 8].) In discussing static prices, Jennings argued that the question of why water (like air and sunlight) had no value was ambiguous because 'as a whole, light and water are of extreme value'. However, any 'definite quantity' could have no value either because there was an 'indefinite quantity…in proportion to the number and to the finite capacities of mankind' or because 'the whole existing quantity cannot be appropriated'. By contrast in 'peculiar circumstances', such as a desert, water could have value (Jennings 1855:210). Diamonds, on the other hand, were 'the most valuable of familiar substances' because when some 'classes of society' were able to consume an abundance of necessities, the 'extreme rarity…[of diamonds] renders them the most prized' of luxury commodities (*ibid.:* 212).

The analytical basis for the marginalist explanation of water and diamond prices was thus a product of the mid-nineteenth century. Jevons was to repeat Jenning's point about the ambiguity of 'value' (Jevons 1878a: 96), having previously formalised and amplified his argument in TPE with the 'fluxional' calculus (as Jennings suggested [1855:35, 260]) and geometry, which clarified the distinction between total and marginal utility. There was, however, no mention of Smith's paragraph or any suggestion of a paradox in the *Natural Elements*. That depiction of Smith came in 1876 when, discussing the distinction between VIU and VIE in a lecture at Owens College, Jevons was reported as telling his students that Smith 'remarked that the most difficult paradox of the whole subject was:

> things which have the greatest value in use have little or no value in exchange. I have no doubt that remark contains the fundamental difficulty in political economy, because…value in one sense can be founded upon use…or what is the same, utility.
>
> (Black 1972 – 81:VI, 79 – 80)

Clearly, the paradox was produced by reading Smith's proposition within the terms of the marginalist theory. Since the latter was the 'true theory of Economy'

(Jevons 1875:6), it was a logical deduction that Smith's statement appeared paradoxical.[24] The transference of authorship to Smith was short-lived, however, as Jevons made no such attribution in his primer where he simply referred to the 'apparent paradox' as a logical problem to be solved (Jevons 1878a:96). These seem to be the only references Jevons made to a paradox. It played no role, for example, in the second edition of TPE when Jevons added a long section expanding his argument that Smith had failed to distinguish between total and marginal utility although he 'evidently means by value in use, the *total utility of a substance*' (Jevons 1970:129 OE).[25] The context of the first appearance of the paradox and its subsequent restriction to the primer suggest that it was formulated as a pedagogic device to illustrate the utility theory.

If that device stabilised the meaning and significance of Smith's paragraph for later marginalist theory, Jevons subsequently proceeded to undermine it. In the *Principles,* where he again discussed Smith's paragraph, Jevons asserted that economists 'as a general rule have acquiesced in this view without further inquiry; hence most of their subsequent perplexities in the theory of value' (Jevons 1905:61). Here and in TPE (Jevons 1871:62), perplexity did not mean that economists were puzzled by Smith's statement, but rather that their arguments were confusing and confused. Indeed, as Jevons noted elsewhere, the perplexity had begun with Smith: 'he should have said "Nothing is more useful than water when water is much wanted." He could not possibly have meant that nothing was more useful than water in a wet mine, or a flooded meadow, or a leaky ship'.[26] The previous pictureof Smith now began to fade and was then effectively obliterated. As Jevons acknowledged in the *Principles,* it made no sense in marginalist theory to say that diamonds could have no VIU but a high VIE:

> as regards a diamond, it is not correct to say that, though a great quantity of other goods may frequently be had in exchange for it, the diamond has scarce any value in use. There is a confusion of ideas here between use in the common everyday meaning and utility in the economic sense. No one would call a diamond ring a very useful thing, but it is generally considered very ornamental, and it, therefore, has utility.
>
> (Jevons 1905:62)

Smith's discussion (referred to above) of how the demand for gold and silver 'arises partly from their utility, and partly from their beauty' (Smith 1976: 189) indicated that he had used utility in the common, everyday meaning. Hence his diamonds and water paragraph 'refers to very obvious ideas' which were not 'analysed as closely as is requisite for a theory of value' (Jevons 1905:62). In Jevons' last words on the topic, the figure of Adam Smith as the author of a paradox which could be solved by the marginalist theory had virtually disappeared. In its place stood the author whose lack of terminological precision

had worried, puzzled and even outraged other critics of the Classical definition of utility.

CONCLUSIONS

For nineteenth-century British political economy, the water and diamonds paragraph in the *Wealth of Nations* was a contested illustrative device which provided a focal point for arguments about the nature of price formation and the meaning and role of utility. The paragraph was thus read and its significance construed in different ways: as a valid statement about the difference between VIU and VIE which required that some care be taken with the use of the word 'value'; as an incoherent, moralistic melange; and as a paradox which, with some clarification, could be explained by the correct theory. The paradox, however, did not exist before Jevons' *Theory* and his primer.

Jevons' reading of Smith as making a distinction between total utility (VIU) and VIE had considerable influence, and it was soon reproduced in the Marshalls' *Economics of Industry* (Marshall and Marshall 1879:68). Here, however, there was no mention of a paradox. The paradox also had no place in Marshall's *Principles of Economics,* where he simply referred to Smith's distinction between VIU and VIE with the laconic comment that 'experience has shown that it is not well to use the word' value to mean theutility of a commodity. It should be used only to mean VIE, as a 'relative term which expresses the relation between two things at a particular time and place' (Marshall 1961:I, 61, see also 129 – 30). The task of depicting the marginalist solution to the classical paradox of value as a defining moment in the history of economics was thus left to subsequent historians and textbooks. Their contribution, in adding the myth that before 1871 economists were unable to explain the relative prices of water and diamonds, was to create a smooth historical narrative in which there was no place for the figure of Jevons, who acknowledged that the marginalist paradox could not be reconciled with its supposed point of origin. Ironic perhaps, but no paradox.

ACKNOWLEDGEMENT

While the usual caveat applies, I would like to thank Peter Groenewegen for helpful suggestions.

NOTES

1 Also known as the Classical paradox of value: why is it that 'water, which has so much value in use, has no value in exchange, while diamonds, which have practically no value in use, are exchanged at high prices[?]' (Ekelund and Hebert 1990: 325).

2 Burtt (1972:53) appears to credit Smith with coining the phrase 'paradox of value'.
3 For a representation of this solution with supply and demand diagrams and a brief history of the paradox in terms of the marginalist denouement, albeit with Alfred Marshall rather than Jevons as hero, see Schmidt 1992. The account in that paper (which has been recommended in the *Journal of Economic Perspectives* as a 'good example for introductory courses of the uses of economic theory' [Saffran 1993: 210]) is consistent with those which, until recently, could be found in textbooks for principles courses in economics where they were used to illustrate the virtues of partial-equilibrium supply and demand analysis (Swales 1993).
4 See, for example, Stephenson 1972 and references cited below.
5 See, for example, the discussion and references in Grotius 1964:351 – 2; Pufendorf 1964:680 – 3.
6 For a survey of the pre-Adamite literature, see Bowley 1973: ch. 2. See also Mandeville 1957:350; Jocelyn 1970:1 – 5; Harris 1757:5; and, for Law, Murphy 1994:30 – 2, 57 – 8 (The relevance of Law in this context was noted by H.D. Macleod [1858:52]). Quesnay also illustrated the distinction between VIU and VIE using diamonds (Meek 1962:89 – 90).
7 The claim that Smith 'had a "blind-spot" with regard to the paradox of value' because, having raised it, he left it 'unresolved' (Swales 1993:237) simply effaces the context of the reference in WN. Cf. Brown (1994:149, 152 – 3, 186 – 8) who uses the differences in the contexts of Smith's references to diamonds and water in the *Lectures* and WN to further clarify why there was no paradox in WN.
8 The analysis appears in a number of places in WN (e.g. Smith 1976:178 – 82, 337 – 8), although its most extensive use was in Book II, especially Chapter IV, to explain the growth of towns and how large landowners lost political and economic power because of their wants for 'the most childish, the meanest, and the most sordid of all vanities' (*ibid.:* 419).
9 For Malthus, natural prices 'depended upon scarcity and difficulty of production'. For this and the agreement with Ricardo on Say, see Malthus 1963:19 – 22, 234 – 5.
10 In a discussion of the Classical paradox unencumbered by textual evidence, it has been claimed that Ricardo 'spent considerable effort trying to understand this seemingly logical inconsistency with relative price determination', eventually resorting to a labour theory of value in arguing that 'water was effectively cheap because it required less labour-effort to acquire than did diamonds' (Schmidt 1992: 1).
11 In a confusing analysis, McCulloch departed from Ricardo in treating labour as 'sweat and toil' and as the only source of wealth (McCulloch 1864:5, 241). Hence he effectively treated labour as akin to a 'real cost' and included 'work done by natural agents as another species of labour' (Bharadwaj 1983:61). See *ibid.:* 60 – 6; De Vivo 1984: ch. 2.
12 OE=original emphasis.
13 Whewell's discussion of the distinction between VIU and VIE (Whewell 1862: 46 – 8) contains four paragraphs, the second of which consists of the familiar quotation from Smith. The following two paragraphs of commentary, from which the quotation above was taken, are also in quotation marks. A search for their source, in the references which Whewell provided at the beginning of his lectures (*ibid.:* ix), proved fruitless. I have thus followed the suggestion of John Pullen that the commentary is actually Whewell's and that, in a work which contains a large

number of extensive quotations from other authors, Whewell made a slip either in writing the text or in correcting the proofs. Pullen suggests that two pieces of evidence support the attribution to Whewell. The first is the relatively modern reference to 'a ventilator, or even a fan' in the commentary (*ibid.:* 47). The second is the frequent use of the first person plural in the two paragraphs, which was characteristic of Whewell's writing (John Pullen, letter to M.White, 31 August 1994).

14 For the relation between labour and VIE, see Whately 1832:152 – 3; on the need to make some distinction between wealth or VIE and use, *ibid.:* 8 – 9.

15 A myth, moreover, which effaces the detailed (non-marginalist) explanations of diamond or pearl prices, based on 'vanity' or 'caprice', provided by John Rae, Senior and Whewell (see White 1992:68 – 9, and the references cited there).

16 Senior claimed that his use of 'utility' was sanctioned by Say and by Malthus (Senior 1965:6; cf. Malthus 1963:234). In view of Malthus' criticism of Say, which was referred to above, this appeal to authority was highly misleading.

17 It has since been claimed that Smith's 'moralistic' definition had changed the 'meaning of the term "use" (or "utility") compared with his predecessors such as Francis Hutcheson' (Hutchison 1978:14; see also Kauder 1953:650). However, on the same page of Hutcheson's *Moral Philosophy,* to which Hutchison refers in support of this claim, it was argued that 'the prices or values in commerce do not at all follow the real use or importance of goods for the support, or natural pleasure of life' (Hutcheson 1969:53).

18 Mill 1909: vol. III, ch. 2; De Quincey 1897:151 – 7.

19 Smith 'only…had to consult in London the New River Board, and to remember that at Paris he had seen it sold retail by those who carried it to houses' (Stark1954: 87). (Cantillon had argued that the price of water in 'the Streets of Paris' was 'the measure of the Labour of the water-carrier' [Cantillon 1964:29].)

20 There is no woman who could not have told him that, in society or at a ball, she who is most richly covered with diamonds overshadows the brilliance of all her rivals. Why should the fair sex not attach importance to them?' (Stark 1954:87).

21 De Quincey suggested that the example showed the distinction between VIU, where a commodity which was regarded as valuable could be 'had for nothing', and VIE, where a commodity was valuable and paid for 'on the spot' (De Quincey 1897:122 – 3).

22 Untitled review in *The Athenaeum,* vol. 36, no. 1845, 7 March 1863:326. H.D. Macleod had also argued that Smith's distinction was 'utterly destructive of all philosophical symmetry. Water is merely the lowest term of a series, of which the highest is a diamond' (Macleod 1858:52). His approach, however, seems more consistent with the Senior strand of the critique which was referred to by Jevons (see below).

23 For references and discussion of this point, see White 1995.

24 When Jevons referred to the paradox in his political economy primer (Jevons 1878a:96) he directed the reader to the discussion of ambiguity in his *Logic* primer (1878b:22 – 6).

25 See Jevons 1970:128 – 31. The only substantive difference with the first edition was the criticism that, in the water and diamonds paragraph, Smith had not made clear that VIE should be depicted as a ratio, which depended on both the relative scarcities of commodities and their respective final degrees of utility.

26 Jevons Archive, John Rylands University Library of Manchester, 6/6/7. The quotation is taken from notes which Jevons began in early 1878 for an edition of WN. The plan was abandoned when H.S.Foxwell informed Jevons that he too was planning an edition of Smith's text. This, however, was never published.

REFERENCES

Bailey, S. (1967) *A Critical Dissertation on the Measures and Causes of Value* (1825), New York: A.M.Kelley (reprint).

Banfield, T.C. (1848) *The Organization of Industry, Explained in a Course of Lectures Delivered at the University of Cambridge in Easter Term 1844,* London: Longman, Brown, Green and Longmans.

Bharadwaj, K. (1978) The subversion of classical analysis: Alfred Marshall's early writings on value', *Cambridge Journal of Economics,* 2 (3):253 – 71.

——(1983) 'Ricardian theory and Ricardianism', *Contributions to Political Economy,* vol. 2:49 – 77.

Black, R.D.C. (ed.) (1972 – 81) *The Papers and Correspondence of William Stanley Jevons,* vols II – VII, London: Macmillan.

——(1987) 'Utility', in J.Eatwell, M.Milgate and P.Newman (eds) *The New Palgrave: A Dictionary of Economics,* vol. 4, Basingstoke: Macmillan (776 – 9).

Bowley, M. (1937) *Nassau Senior and Classical Economics,* London: George Allen and Unwin.

——(1973) *Studies in the History of Economic Theory before 1870,* London: Macmillan.

Brown, V. (1994) *Adam Smith's Discourse: Canonicity, Commerce and Conscience,* London and New York: Routledge.

Burtt, E.J. (1972) *Social Perspectives in the History of Economic Theory,* New York: St Martin's Press.

Cantillon, R. (1964) *Essai sur la Nature du Commerce en General* (1755), translated and edited by Henry Higgs, New York: A.M.Kelley.

Cooter, R. and Rappoport, P. (1984) 'Were the ordinalists wrong about welfare economics?', *Journal of Economic Literature,* 22 (2):507 – 30.

De Marchi, N. (1987) 'Paradoxes and anomalies', in J.Eatwell, M.Milgate and P. Newman (eds) *The New Palgrave: A Dictionary of Economics,* vol. 3, London: Macmillan.

De Quincey, T. (1897) *The Collected Writings of Thomas De Quincey,* vol. IX, edited by David Masson, Soho: Black.

De Vivo, G. (1981) 'John Stuart Mill on value', *Cambridge Journal of Economics,* 5 (1): 67 – 9.

——(1984) *Ricardo and His Critics: A Study of Classical Theories of Value and Distribution,* Studi e Ricerche dell'Istituto Economico, no. 23, Facoltà di Economia e Commercio, Università degli Studi di Modena.

Edgeworth, F.Y. (1987) 'De Quincey, Thomas (1785 – 1859)' (1894), in J.Eatwell, M. Milgate and P.Newman (eds) *The New Palgrave: A Dictionary of Economics,* vol. 1, London: Macmillan.

Ekelund, R.B. Jr. and Hebert, R.F. (1990) *A History of Economic Theory and Method,* 3rd edn, New York: McGraw-Hill.

Galbraith, J.K. (1989) *A History of Economics,* Harmondsworth: Penguin.

Groenewegen, P. (1982) 'Thomas de Quincey: "Faithful Disciple of Ricardo"?', *Contributions to Political Economy,* vol. 1:51 – 8.

Grotius, H. (1964) *De Jure Belli ac Pacis Libri Tres* (1625), vol. II, translated by Francis W.Kelsey, New York/London: Oceana/Wiley.

Harris, J. (1970) *An Essay Upon Money and Coins. Part I* (1757), Wakefield: S.R. Publishers (reprint).

Hutcheson, F. (1969) *A System of Moral Philosophy* (1755), vol. II, Hildesheim: Georg Olms.

Hutchison, T.W. (1956) 'Bentham as an economist', *Economic Journal,* 66, June: 288 – 306.

——(1978) *On Revolutions and Progress in Economic Knowledge,* Cambridge: Cambridge University Press.

Jevons, W.S. (1871) *The Theory of Political Economy,* London: Macmillan.

——(1875) 'The Progress of the mathematical theory of political economy with an explanation of the principles of the theory', *Transactions of the Manchester Statistical Society, Session 1874 – 5,* 1 – 19.

——(1878a) *Political Economy,* London: Macmillan.

——(1878b) *Logic,* 2nd edn, London: Macmillan.

——(1905) *The Principles of Economics. A Fragment of a Treatise on the Industrial Mechanism of Society and Other Papers,* London: Macmillan.

——(1970) *The Theory of Political Economy,* edited by R.D.Collison Black, Harmondsworth: Pelican.

Jocelyn, J. (1970) *An Essay on Money and Bullion* (1718), Wakefield/New York: S.R. Publishers/Johnson Reprint Corporation.

Kauder, E. (1953) 'Genesis of the marginal utility theory', *Economic Journal,* 63, no. 251, September: 638 – 50.

Kaushil, S. (1973) 'The case of Adam Smith's value analysis', *Oxford Economic Papers,* 25 (1), March: 60 – 71.

McCulloch, J.R. (1864) *The Principles of Political Economy. With Some Inquiries Respecting their Application,* 5th edn, Edinburgh: Black.

Macleod, H.D. (1858) *The Elements of Political Economy,* London: Longman, Brown, Green, Longmans and Roberts.

Malthus, T.R. (1963) *Definitions in Political Economy* (1827), New York: A.M. Kelley (reprint).

Mandeville, B. (1957) *The Fable of the Bees,* edited by F.B.Kaye, Oxford: Clarendon Press.

Marshall, A. (1961) *Principles of Economics,* ninth variorum edition, edited by C.W. Guillebaud, Macmillan, London.

Marshall, Alfred and Marshall, Mary Paley (1879) *The Economics of Industry,* London: Macmillan.

Meek, R.L. (1962) *The Economics of Physiocracy. Essays and Translations,* London: Allen and Unwin.

Mill, J.S. (1909) *Principles of Political Economy,* edited by W.J.Ashley, London: Longman, Green and Co.

Murphy, A. (ed.) (1994) *John Law's 'Essay on a land bank',* Dublin: Aeon Publishing.

Oser, J. and Brue, S.L. (1988) *The Evolution of Economic Thought,* 4th edn, San Diego CA: Harcourt Brace Jovanovich.

Pufendorf, S. (1964) *De Jure et Gentium Libri Octo* (1688), vol. II, translated by C. H.Oldfather and W.A.Oldfather, New York/London: Oceana/Wiley.

Ricardo, D. (195la) *On the Principles of Political Economy,* in *Collected Works and Correspondence of David Ricardo,* vol. I, edited by Piero Sraffa with the collaboration of M.H.Dobb, Cambridge: Cambridge University Press.

——(1951b) 'On protection to agriculture' (1822), in *Collected Works and Correspondence of David Ricardo,* vol. IV.

Roll, E. (1966) *A History of Economic Thought,* 2nd edn, London: Faber.

Saffran, B. (1993) 'Recommendations for further reading', *Journal of Economic Perspectives,* 7 (2):207 – 12.

Say, J.B. (1964) *A Treatise on Political Economy, or the Production, Distribution and Consumption of Wealth* (1821), New York: A.M.Kelley (reprint).

Schabas, M. (1990) *A World Ruled by Number. William Stanley Jevons and the Rise of Mathematical Economics,* Princeton NJ: Prince ton University Press.

Schmidt, R.H. (1992) 'Diamonds and water: a paradox revisited', *FRBSF Newsletter ,* Federal Reserve Bank of San Francisco, no. 92 – 43, 4 December: 4pp.

Schumpeter, J.A. (1963) *History of Economic Analysis* (1954), edited by Elizabeth Boody Schumpeter, London: George Allen and Unwin.

Senior, N. (1965) *An Outline of the Science of Political Economy* (1836), New York: A.M.Kelley.

Smith, A. (1976) *An Inquiry into the Nature and Causes of the Wealth of Nations,* edited by R.H.Campbell, A.S.Skinner and W.B.Todd, vol. I, Oxford: Oxford University Press.

——(1978) *Lectures on Jurisprudence,* edited by R.L.Meek, D.D.Raphael and G. Stein, Oxford: Clarendon Press.

Stark, W. (1954) *Jeremy Bentham's Economic Writings,* critical edition based on his printed works and unprinted manuscripts, vol. 3, London: Allen and Unwin.

Stephenson. M.A. (1972) 'The paradox of value: a suggested interpretation', *History of Political Economy,*4 (1):127 – 39.

Swales, J. (1993) 'The paradox of value: six treatments in search of the reader', in Willie Henderson, Tony Dudley-Evans and Roger Backhouse (eds) *Economics and Language,* London and New York: Routledge.

Twiss, T. (1847) *View of the Progress of Political Economy in Europe since the Sixteenth Century,* London: Longman, Brown, Green and Longmans.

Whately, R. (1832) *Introductory Lectures on Political Economy, Delivered in the Easter Term, MDCCCXXI,* 2nd edn, London: Fellowes.

——(1853) *Easy Lessons on Money Matters for the Use of Young People,* 13th edn, London: Parker.

Whewell, W. (1862) *Six Lectures on Political Economy Delivered at Cambridge in Michaelmas Term, 1861,* Cambridge: Cambridge University Press.

——(1971) 'Mathematical exposition of some doctrines of political economy' (1850), reprinted in W.Whewell, *Mathematical Exposition of Some Doctrines of Political Economy (1829, 1831 and 1850),* New York: A.M.Kelley.

Whitaker, J.K. (ed.) (1975) *The Early Economic Writings of Alfred Marshall,* vol. 1, London: Macmillan.

White, M.V. (1992) 'Diamonds are forever(?): Nassau senior and utility theory', *The Manchester School,* 60 (1):64 – 78.

——(1994) 'The moment of Richard Jennings: the production of Jevons's marginalist economic agent', in Philip Mirowski (ed.) *Natural Images in Economics. Markets Read in Tooth and Claw,* New York: Cambridge University Press.

——(1995) ' "No Matter of Regret": the Cambridge critique(s) of Jevons' hedonics', forthcoming.

2
THE RICARDO-MALTHUS DEBATE
Effective demand, technical change and the limits to growth

John F.Berdell

CONTENDING ANALYSES OF TECHNICAL CHANGE, DISTRIBUTION, AND DEMAND

Only an especially brave student would seek to delimit the boundaries of G. C.Harcourt's writings. His early theoretical contributions disregard the borderline between accounting and economics (one Hicks trespassed as well), while later contributions often surfaced quietly within interpretations and reviews. Not only is there a fluid movement between interpretation and theoretical extension, but across disciplinary dividing lines as well. Indeed the whole broad yet fast moving post-Keynesian stream, with which his work is so closely associated, has rarely paused for such boundaries.

While the boundaries to this watershed are difficult to locate the Classical economists have always figured prominently therein. Few elements of the Classical literature can have gained more attention than the Malthus-Ricardo debate over the measure, causes, and nature of economic growth. Malthus' thoughts on the interaction between economic and demographic growth have of course also provided considerable stimulus to economists and demographers well outside of the post-Keynesian tradition. This chapter develops a non-linear dynamic model of economic and demographic growth which aims to embody the most essential elements of Malthus's policy-oriented analysis of advanced economies. The model and its results are then compared with a very simple model of Ricardian growth so as to contrast their approaches to economic dynamics, and identify the basis of their competing policy prescriptions. While it is not within the scope of this paper to trace the later fate of these arguments, it will be clear to many readers that they came to trouble and involve not only Geoff Harcourt, but almost all of the twentieth-century growth theorists who have figured prominently in his writings.

I draw several elements from the wealth of existing Malthus models and place them within a non-linear continuous-time dynamic model in which the accumulation of capital, the growth of population, the distribution of income, and technical change participate in a process of circular and cumulative causation in

which trend and cycle are inseparable. Each element of the model is developed and discussed individually before their joint action across time is examined. Capital accumulation is developed in the next section, which presents a formulation of Malthus's ever-contentious principle of effective demand. This formulation follows a line of interpretation and modelling in which the dispute between Malthus and Ricardo is seen as one over the determinants of the *rate of growth* of output and employment at any moment in time, rather than with their *level* at that moment in time. The formulation of the principle of effective demand used here renders the profit rate and rate of economic growth as functions of class-specific spending behaviour, the distribution of income and independently specified investment behaviour. To highlight the controversy over the rate of growth of output, a common production function relating the level of output to inputs is specified for the Ricardian and Malthusian models. The controversy is not only over the role of effective demand as a determinant of the profit rate and distribution which figures so prominently in Harcourt (1988) and Kaldor (1955 – 6). It also concerns at least some of the issues which Harcourt (1986) found to be at stake in recent debates on the role and implications of resource constraints and scarcity as limits to growth.

The third section argues that this approach to the principle of effective demand is potentially consistent with seemingly disparate interpretative approaches: those stressing the satiation of demand, as well as those stressing the ratio of productive to unproductive labour. The model also provides a basis for assessing the similarities and differences between

Malthus' and Keynesian analyses of macro dynamics and growth. Malthus' principle of population did not constitute a point of contention between the authors, hence the fourth section introduces a common specification for demographic behaviour which closes the system and allows us to contrast the movement of the Ricardian and Malthusian economic-demographic system over time.

Malthus' final position on the relationship between population growth and technical change in agriculture is introduced in the fifth section. Technical change generates a pattern of cyclical growth broadly consonant with his discussion of economic oscillations and the concatenation of economic causes and effects. The pattern of growth generated by the endogenous growth of population, capital and technology is one in which trend and cycle are conceptually inseparable. Causes become consequences as the variables enter into a process of circular and cumulative causation. The evolution of the system is rendered path dependent—the future evolution of the system is dependent upon its past history—by the non-linearityand non-reversibility of social and technical change. It is the activity of this non-linearity which inhibits the application of simple analytical methods and renders the use of simulation techniques attractive.

THE PRINCIPLE OF EFFECTIVE DEMAND

Because Malthus' principle has been the subject of continuous interpretation and formalisation, the impression has gained ground that economists can not agree on the basic nature of his principle or its relationship to other theories of growth. Yet in the wake of Keynes' laudatory account of the first Cambridge economist, a series of authors concurred in regarding its operative feature as the interaction of class-specific spending and saving behaviour with investment behaviour which is governed by (at least partially) independent considerations. This section briefly develops a formulation of this approach which is appropriate to a continuous-time model.

Following O'Leary (1942, 1943), Link (1959), Vatter (1959), Mishra (1978), Eltis (1980, 1984) and Costabile and Rowthorn (1984, 1985), we will view the principle as explicable in terms of the interaction between income-group specific propensities to save and the rate of investment undertaken at each rate of profit. It is useful to divide this interpretative approach into two branches. After the *General Theory* it was natural to seek an equilibrium theory of output in Malthus which could explain the failure of market forces to eliminate unemployment. This is what Link and O'Leary sought in Malthus in the 1940s and early 1950s.[1] Subsequent authors, however, have interpreted the discussion between Ricardo and Malthus as one primarily concerned with the determinants of the *rate* of economic *growth,* in which special attention is given to the current profit rate as the principle determinant of the growth in the demand for labour. Here saving and investment behaviour determine the rate of growth of output at a moment in time rather than its level. These models are of the type which Kaldor named 'Keynesian theories of distribution'. It is along the lines of this second approach that we proceed here, taking the level of output and employment at a moment in time as given. Although an assessment of the relationship to Keynes is impossible here, we can note in passing the absence of a highly developed view of uncertainty and Harcourt's (1985:9) view that 'Keynes' theory of effective demand without expectations and uncertainty is like Hamlet without the Prince'.[2]

Malthus had already been writing about the growth of human population for two decades by the time he examined the causes of stagnation in his *Principles*. The influence of his population theory on his theory of capital accumulation was evidently quite strong.[3] For Malthus the growth of a human population is related to the difference between the current standard of living and a minimally acceptable standard of living. So by analogy the growth of capital was governed by the difference between the current profit rate and a minimally acceptable profit rate (1989a:370). Unlike the minimal standard of living, changes in the minimally acceptable rate of profits π_{min} on the part of investors does not receive much attention in Malthus' analysis, and so will be taken as constant. If we follow this line of thought investment, expressed as the growth rate of the capital stock, is a function of the current profit rate:

$$I / K = f(\pi - \pi_{min}) f'(x) > 0, f(0) = 0 \tag{1}$$

If the minimally acceptable profit rate is constant over time and if a linear function is acceptable we may write:[4]

$$I / K = i_0 + i_1 \bullet \pi, i_0 < 0, i_1 > 0 \tag{2}$$

Saving behaviour is group-specific. Merchants and manufacturers constitute a single group and have a higher propensity to save than do landlords. Labourers do not save. Saving as a fraction of national income is consequently a function of the fraction of national income which accrues to merchants, manufacturers and landlords (1989a:466). Let represent total profits, *R* total rents, and *Y* output (all in terms of the one 'real' good).

$$\begin{aligned} &S / Y = s_0 + s_1 \bullet (\Pi / Y) + s_2 \bullet (R / Y) \\ &s_0 < 0 < s_1, s_2 < 1 \end{aligned} \tag{3}$$

Habits and propensities—two terms often employed by Malthus—provide a specification of saving behaviour which is independent of investment behaviour. It is this independence which allows one to look for the conditions under which saving equals investment. Indeed the equality of saving and investment will be used, following Kaldor, to determine the distribution of the product between wages and profit.

Both Malthus and Ricardo at times appealed to rent theory for a purely technological explanation of the share of rent in output, an appeal easily granted only in the case of a one good, or corn, model in which labour and land are the only productive factors, and capital is simply the wage bill. The diminishing productivity of successive units of labour used in agriculture is based on the diminishing fertility of new land taken under cultivation and the diminishing result of applying labour to land already under cultivation. Diminishing returns determines the rent R^{eq} which accrues to landlords if all agricultural tenures are perfectly adjusted to the current level of output. (1989a:134 – 5, 153 – 4). To facilitate comparison of Malthus' and Ricardo's results it is convenient to utilise a common specification for corn production *Y* as a (diminishing) function of labour employed *E*.

$$Y = \alpha \bullet E - \beta \bullet E^2 \tag{4}$$

so that the marginal product *dY/dE* and equilibrium rent R^{eq} are:

$$dY / dE = \alpha - 2 \bullet \beta \bullet E \tag{5}$$

$$R^{eq} = (Y / E) - (dY / dE) = \beta \bullet E^2 \tag{6}$$

If output and employment are taken as given then rent is technologically given and our problem is only how to divide the product net of rent between wages and

profit. This is done by setting saving equal to investment where we substitute for π according to:[5]

$$\pi = ((dY / dE) - C) / C \tag{7}$$

so we have:

$$R^{eq} = (Y / E) - (dY / dE) = \beta \bullet E^2 \tag{8}$$

or

$$\pi = \frac{(i_0 + s_1) \bullet (dY / dE) \bullet E - s_0 \bullet Y - s_1 \bullet (dY / dE) \bullet E - s_2 \bullet R}{s_0 \bullet Y + s_1 \bullet (dY / dE) \bullet E + s_2 \bullet R - i_1 \bullet (dY / dE) \bullet E} \tag{9}$$

Saving propensities and technology determine the distribution of output and its rate of growth, since the profit rate determines the rate of investment. The rate of profit determined here is given the rubric 'the regulating principle of profits' in the second edition of Malthus' *Principles*. The limiting principle of profits' refers to Ricardo's determination of the profit rate by the difference between real wage and the marginal product of labour.[6] (Ricardo's formulation will be explicitly contrasted to Malthus' in what follows.) Malthus fully agreed with Ricardo that diminishing returns in agriculture could eventually constrain the profit rate, but in the 'actual state of things' it was the regulating principle that mattered and should form the basis for policy conclusions (1989a:313 – 26).[7] Note that an increase in the propensity to save on the part of either of the saving classes reduces the profit rate, and consequently the rate of investment.[8] Hence Malthus' warnings that a sudden conversion of consumption flows into saving could stifle growth. It was growth in the demand for labour which particularly concerned Malthus as he believed that the proportion of the population in employment was an important determinant of per-capita income. The change in employment is given by Equation (2) in conjunctionwith the real wage given in Equation (8):

$$\dot{E} = I / C \tag{10}$$

This so-far static interpretation of demand and profits in Malthus is consonant with Sowell's (1963) view that the general glut controversy was over the rate of growth and the incentives to further increases in output, rather than the determinants of the current level of output and employment. The relevance to the Corn Law debate is not far over the horizon, as any policy which attempted to suddenly increase profits at the expense of rents could reduce effective demand and induce stagnation.

ALTERNATE FORMULATIONS OF THE PRINCIPLE AND ITS KEYNESIAN CONTENT

While the saving and investment approach to the principle of effective demand has been adopted by a succession of authors, others have claimed that the principle is based on fundamentally different categories and concerns. Spengler (1965) stresses the service sector, and suggests that the ratio of productive to unproductive labour is the principal determinant of the rate of growth. Materially unproductive (i.e. service) labour plays 'a double role in Malthus' model'. They refrain from producing material goods while providing extra demand for them. They stop the prices of material goods from dropping towards, or falling below their costs of production. The crucial issue is the same as the saving and investment approach—the determination of the price level and rate of profit. Since landlords have a much higher propensity to consume services than do capitalists, the optimal ratio of productive to unproductive labour can be regarded as merely a transformation of the ratio of profits to rents which yields the highest profit rate in Equation (10).

Rashid (1977) presents a stronger challenge to the saving and investment approach which, he maintains, fails to provide a satisfactory treatment of Malthus' view of gluts. Rashid argues that satiation is the primary cause of short-run downturns in profits and trade for not only Malthus, but for an impressively comprehensive list of 'dissenters from Say's law'. His emphasis is on the necessity and difficulty of forming new habits of consumption among capitalists and landlords. Without new consumption habits their desire to secure the economic position and prestige of their families might lead them to save a growing portion of income and so constrain demand. This is reminiscent of Duesenberry's (1949) work on the foundations of household consumption behaviour, which was in retrospect generally taken as providing a sociological, if somewhat critical, foundation for the standard Keynesian consumption function. Malthus' attention to the slow development of habits and wants might similarly justify a consumption-function approach to modelling his 'regulating principle'.

The principal objection to Keynesian or pseudo-Keynesian interpretations of Malthus must be that Malthus simply did not distinguish saving from investment.[9] Certainly we can find passages in which saving is not clearly distinguished from investment. It is important to recall, however, that in the model presented here saving is always equal to investment because the profit rate is always at the (instantaneous) equilibrium value which equates them. The instantaneous equilibrium assumption of the model implies that the independence of saving and investment behaviour is not reliant upon hoarding and a distinction between ex-ante and ex-post investment. Many interpretations of Malthus' principle of effective demand are not compatible with the saving and investment approach. Approaches can easily emphasise different passages and causal linkages, so as to suggest that the saving or investment schedules used here shift sufficiently erratically as to render them uninformative. The large

potential for reconciling seemingly disparate interpretations of Malthus' principle should, however, indicate that there is greater room for agreement than the tone of the literature suggests.

Since it was Keynes' laudatory biographical essay which inspired the postwar effort to formalise Malthus' principle of effective demand, it is important to situate that principle against Keynes' writings. As formulated here, Malthus' model is 'Keynesian' but it is considerably closer to the Keynes of the *Treatise on Money* than of *The General Theory* or after. We have implicitly posited a flex-price model in which price adjusts instantly and quantity adjusts only across time, a procedure which, despite the absence of inventories, resembles the method Amadeo (1989:22) attributes to the *Treatise On Money*. The model is of the type used by Kaldor (1980) in his work on growth and distribution, and in his interpretation of Keynes' *Treatise on Money* (1980:209, 227). As Malthus neither sought to construct an equilibrium theory of output at an instant of time (nor to attribute to the money market a fundamental role in determining the economy's position or trajectory) it is with good reason that it is usually observed that Keynes was too generous when he suggested that the basis of *The General Theory* could be found in Malthus. For some readers the greatest difference between the Keynesian and Malthusian principles of effective demand may lie in the role allocated to uncertainty. In contrast to the essential role which Harcourt and O'Shaughnessy (1985:9) have ascribed to uncertainty in the *General Theory,* it plays no role in this interpretation of Malthus' effective demand. Once again, *if* there is a debt it is more likely to be found in the *Treatise*.[10]

THE PRINCIPLE OF POPULATION

In contrast to the great difference separating Malthus' and Ricardo's *Principles* on the question of distribution, their treatment of population is remarkably similar once semantic differences are put to one side. This is of course no surprise, as it has long been accepted that Ricardo and his followers adopted Malthus' demographic theory. The treatment of Malthus' 'final' position on population dynamics adopted here builds on a number of models in which population growth is specified as a relatively simple function of current income and what is somewhat confusingly termed a 'subsistence' level of income.[11] It must, however, be emphasised that in Malthus' late writings it is an 'optimistic' theory in which standard of living can secularly rise. In this model of population dynamics, no attention is given to the demographics of capitalists and landlords; the population in question is composed entirely of labouring families. In order to simplify exposition, the standard of living will be taken as identical to per capita income among the labouring population. (Though Malthus felt that factors such as urbanisation significantly reduced the quality of life.) Thus simplified, per capita income *pci* among the labouring class depends upon only two variables: the real wage *C* and the proportion of the population, in employment *E/P*.

$$pci = C \bullet E / P \tag{11}$$

Once Malthus' use of the concept of a 'standard of wretchedness' in place of Ricardo's 'natural rate of wages' is accounted for, the structure of their demographic theory appears substantively the same. For both authors, human populations grow when they experience a higher standard of living than that which their culture, as embodied in habits and customs, has historically come to regard as the minimally acceptable standard of living. Malthus called this minimum the 'standard of wretchedness', while Ricardo called its value expression the 'natural price of labour'. Unlike Ricardo, Malthus gave considerable attention to variations in the labour participation ratio. Women and children were more likely to be employed in expansionary periods than in recessions or depressions, hence family income was strongly pro-cyclical. Malthus assumed that greater demand for women's labour raises fertility. On the basis of these assumptions, population growth can be given as a simple function of the gap between per capita income and the standard of wretchedness *S*.

$$\dot{P} = \gamma \bullet (pci - S) \bullet P, \gamma > 0 \tag{12}$$

Both Malthus and Ricardo thought that the standard of wretchedness had increased across time, and were in favour of schemes designed toencourage the lower classes to regard more 'luxuries' as 'necessities', that is, as part of the standard of wretchedness. While their interest in education clearly shows that non-economic variables were thought to play a considerable part in modifying the cultural attributes which constituted the 'standard of wretchedness', a simple specification which ignores these complexities can still provide considerable insight. Simplifying matters considerably then, the historical evolution of the standard of wretchedness is determined by discrepancies between the current level of income and the current standard of wretchedness.[12]

$$\dot{S} = \delta \bullet (pci - S) \quad \text{if} \quad pci > S$$

$$\dot{S} = \delta' \bullet (pci - S) \quad \text{if} \quad pci < S,\ \delta > \delta' > 0 \tag{13}$$

Equations (12) and (13) constitute a summary demographic theory which can be used to interpret Ricardo and Malthus.[13] Together, the principles of effective demand and population constitute a formalisation of Malthusian growth theory in which the accumulation of capital and growth of population interact, technology remaining constant. It is only after introducing technical change, however, that we can arrive at the author's differences on the politics and policies governing the limits to growth.

TECHNICAL CHANGE, AGRICULTURAL IMPROVEMENT, AND LONG-RUN GROWTH

The second and subsequent editions of the 'Essay on population' contain a remarkably optimistic assessment of the future course of the standard of living, though often overshadowed by the dire consequences which might follow from policy errors (1986: vol. 3, 448 – 56; 1989b:229). Agricultural improvement plays a crucial role in his expectation that per capita income will secularly rise in Britain. The secular trend is the result of a cyclical process in which the growth of agricultural output and the growth of population alternatively take precedence, so that per capita income alternately rises and falls. It is the standard of wretchedness which generates a secular trend out of this cyclical action: during any period in which per capita income rises, the standard of wretchedness rises and presents a limitation on the subsequent reductions in the standard of living.

For Malthus, profitability is the proximate determinant of the rate of agricultural technical change. An increase in demand for agricultural products, or a fall in agricultural costs of production, will increase agricultural profits and both stimulate and finance the introduction of 'improvements' which are capital-embodied technical changes. Malthus insists that landlords must refrain from stifling this process by prematurely raising rents when profits are rising; to do so would deprive the farmer of the funds for investment. Landlords should encourage technical change because, *contra* Ricardo, it raises their income. Here we assume that the aggregate rent charged by landlords R lags behind the maximum they could at any moment demand R^{eq}.

$$\dot{R} = \zeta \bullet (R^{eq} - R) \quad \zeta > 0 \tag{14}$$

I will examine reductions in the intercept parameter of the agricultural production function which correspond to Malthus' and Ricardo's differing positions on the movement of rent (both in output and as a share) over time. In keeping with Malthus' discussion, technical changes, an essential element of improvements, are made available to farmers when profitability rises. More particularly, what we may call the 'potential' slope moves towards vertical when the profit rate increases. The rate at which the actual slope changes is determined by the rate of investment. (Since labour is the only input to agriculture it seems appropriate to make improvement a function of 'real investment', i.e. the flow of new labour into agriculture.)

$$\dot{\beta} = \eta \bullet (I / C) \bullet 1 - \beta / \beta^{Potential}) \bullet \beta \quad \text{if} \quad I > 0, (\eta > 0) \tag{15}$$

$$\dot{\beta} = 0 \quad \text{if} \quad I < 0$$

The 'potential' slope in turn shifts according to:

$$\Delta\beta^{Potential} = -\theta \bullet \beta \quad \text{iff} \quad \pi > 0 \; \dot{\pi} > 0 \; t^e > t^d, \; 0 < \theta < 1 \tag{16}$$

where t^e is the time elapsed since the last technical change, t^d is the minimum delay between technical changes and θ is the amounts by which the potential slope shifts.[14] In the following simulations the slope parameter is reduced by 5 per cent every fifty periods if the profit rate is positive and rising.

CAPITAL ACCUMULATION AND POPULATION GROWTH OVER TIME

We are now in a position to assess the differences between the Malthusian and Ricardian approaches to economic and demographic growth. Table 2.1 summarises the Malthus system described above and Table 2.2 presents a Ricardian system which is specified to facilitate comparison. Saving and investment behaviour are not distinguished in the Ricardo model so that the rate of growth of the capital stock is a linear function of the rate of profit. The real wage is simply the capital stock divided by the labour force. Employment in the Ricardian system is simply half the size of the popula tion , hence its growth is directly determined by population (the supply side) rather than the Malthusian employment equation (the demand side). The Ricardian real wage can be taken as given at an instant in time, because the wage fund is given by the past history of investment while the labour force is given by demographic history. Hence the behaviour of the Ricardian real wage over time is quite straightforward. It rises if the capital stock grows more quickly than the labour force and vice-versa.

The movement of the principal macroeconomic variables of these systems over time, from similar initial conditions and with similar parameters, may be ascertained with the aid of numerical simulation and is contrasted in Figures 2.1, 2.2 and 2.3. Figure 2.1a illustrates the movement of the real wage and the marginal product of labour in the Malthusian and Ricardian systems as employment increases. The Malthusian marginal product and real wage both decline at first, under the joint influence of diminishing returns to scale and the determination of distribution by the principle of effective demand. Along this declining path the profit rate steadily declines, until profits fall below the minimum profit rate. At this point investment is negative and the real wage and employment fall, sending profits in the other direction. The first ‘wave’ of technical change is triggered by the rising profit rate and the production function is potentially able to shift outward. Productivity will only increase, however,

Table 2.1 The Malthus system

$Y = \alpha \cdot E - \beta \cdot E2$	$\alpha = 18.5$, $\beta = 0.05$, $E = 100$ at $t = 0$	
$dY/dE = \alpha - 2 \cdot \beta \cdot E$		
$R^{eq} = \beta \cdot E^2$		
$C = \dfrac{(s_0 i_1 Y + s_1 \cdot (dY/dE) \cdot E + s_2 \cdot R - i_1 \cdot (dY/dE) \cdot E)}{((s_1 + i_0 - i_1) \cdot E)}$	$i_0 = -0.0002$, $i_1 = 0.22$	$s_0 = -0.05$, $s_1 = 0.9$, $s_2 = 0.1$
$\pi = ((dY/dE) - C)/C$		
$I = (i_0 + i_1 \cdot \pi) \cdot C \cdot E$		
$dE/dt = I/C$		
$pci = C \cdot E/P$		
$dP/dt = \gamma \cdot (pci - S) \cdot P$		$\gamma = 0.02$
$dS/dt = \delta \cdot (pci - S)$	for $pci > S$	$\delta = 0.06$
$dS/dt = \delta' \cdot (pci - S)$	for $pci < S$	$\delta' = 0.03$
$dR/dt = \zeta \cdot (R^{eq} - R)$		$\zeta = 0.16$
$d\beta/dt = \eta \cdot I / C \cdot (1 - \beta/\beta^{Potential}) \cdot \beta$	if $I > 0$	$\eta = 2$
$d\beta/dt = 0$	if $I < 0$	
$\Delta\beta^{Potential} = -\theta \cdot \beta$	iff $\pi > 0$, $d\pi/dt > 0$, $t^e > t^d$	$\theta = 0.1$, $td = 50$

Table 2.2 The Ricardo system

$Y^R = \alpha \cdot E^R - \beta \cdot (E^R)^2$	$\alpha = 18.5$, $\beta = 0.05$ and $E^R = 100$ at $t = 0$	
$dY^R/dE^R = \alpha - 2 \cdot \beta \cdot E^R$		
$R^R = \beta \cdot (E^R)^2$		
$\pi^R = ((dY^R/dE^R) - C^R) / C^R$		
$\Pi^R = \pi^R \cdot C^R \cdot E^R$		
$dK^R/dt = I^R = (i_0 + i_1 \cdot \pi^R) \cdot (C^R \cdot E^R)$	$i_0 = -0.0002$, $i_1 = 0.22$, $K^R = 824$ at $t = 0$	
$C^R = K^R/E^R$		
$E^R = 0.5 \cdot P^R$		
$pci^R = C^R \cdot E^R/P^R$		
$dP^R/dt = \gamma \cdot (pci^R - S^R) \cdot P^R$	$\gamma = 0.02$	
$dS^R/dt = \delta \cdot (pci^R - S^R)$	for $pci^R > S^R$	$\delta = 0.06$
$dS^R/dt = \delta' \cdot (pci^R - S^R)$	for $pci^R < S^R$	$\delta' = 0.03$

Note: β is determined by the Malthus system

if investment is positive, and the rate at which it increases, as we have seen, depends both upon investment and the extent of the recently created potential for productivity growth. With the parameters used in Table 2.1, the process of endogenous productivity growth raises the marginal product of labour over time, as is clearly visible in Figure 2.1a. Several technical changes are clearly visible, and the real wage in both systems is seen to converge to the new (flatter) marginal product schedule after each productivity shock. The Ricardian system exhibits considerably more dramatic real wage growth after each technical shock. Each shock immediately increases profits and investment because profitability is directly determined by the gap between the real wage and the marginal product of labour in the Ricardian system. Investment simply takes the form of a larger wage fund. Hence the real wage rises quickly to the new stationary state. Savings and investment behaviour generate very different behaviour for both profitability and the real wage in the Malthusian system. The increase in the real wage is far less dramatic in the Malthusian system. In contrast to the Ricardian system the downward movement in the wage is as pronounced as its upward movement. Yet as Figure 2.2b shows, the Malthusian system actually generates a higher level of per capita income. The source of higher income in the Malthusian system lies with employment growth combined with slower population growth. Figure 2.1b shows the rising labour participation rate in the Malthusian system. (The labour participation rate is constant by construction in the Ricardian system.)

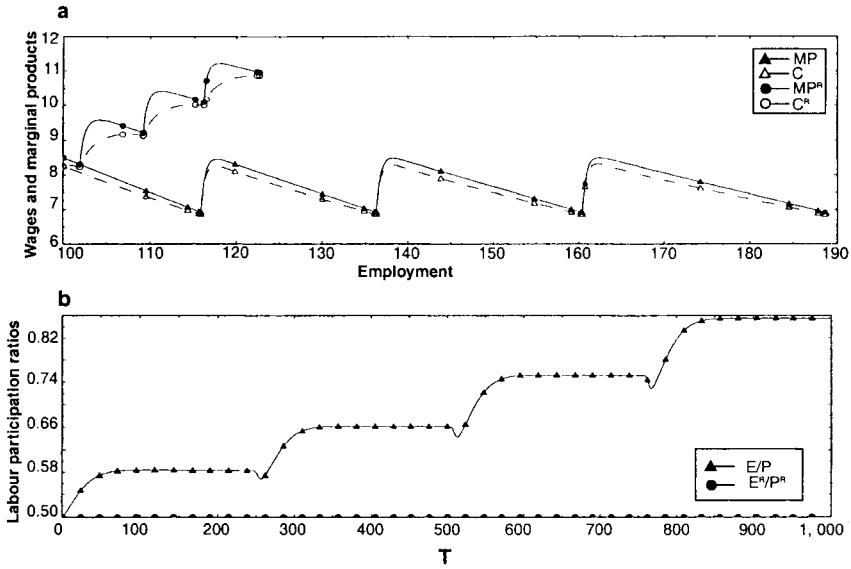

Figure 2.1 Movement of the principal macroeconomic variables of Ricardian and Malthusian systems over time: wages and labour

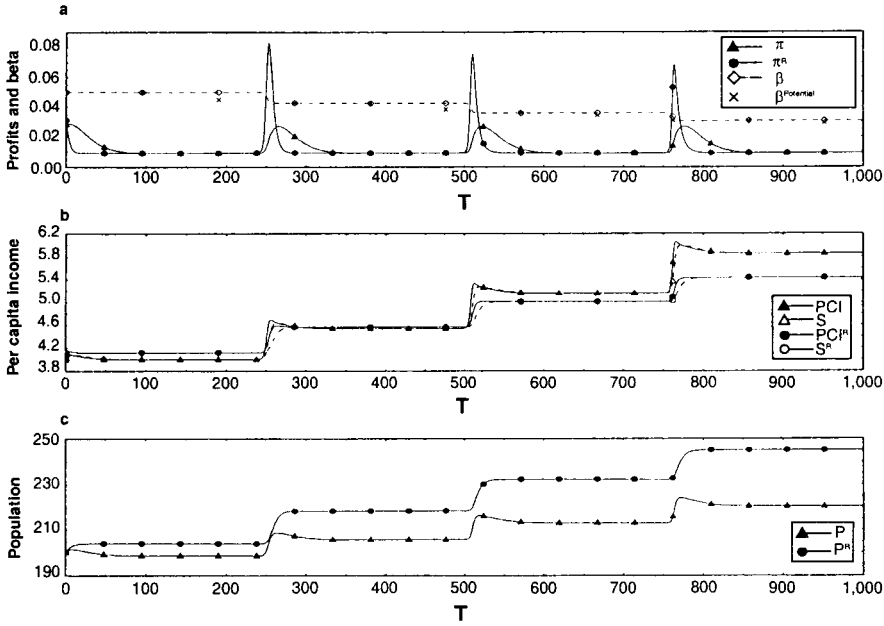

Figure 2.2 Movement of the principal macroeconomic variables of Ricardian and Malthusian systems over time: profits, per capita income and population

Figure 2.2a corroborates that periods of technical change are coterminous with periods of high profits and investment. The slope term falls during periods of high profits and investment, but remains constant in between these periods. The higher rate of profit in the Ricardo system causes capital to accumulate quickly, much more quickly than population and the labour force grow. Hence the real wage in the Ricardian system quickly converges towards the marginal product of labour, and the profit rate falls as it does so. The figure indicates that in the high growth periods following a technical change, the profit rate in the Malthusian system is lower than that in the Ricardian system; this reflects Malthus' proposition that the 'regulating' principle of profits (effective demand) rather than the 'limiting' principle of profits (Ricardo's) was often the binding constraint on growth. The movement of the profit rate and wages in these systems reflects Malthus' opinion that effective demand constrained British economic growth in the 1820s, and Ricardo's contrary opinion that low profits and sluggish growth were caused by diminishing returns in agriculture. While the specification of effective demand used here draws heavily upon Kaldor's, it does not share what Harcourt (1988:105) has termed Kaldor's 'quixotic conviction' that labour is fully employed.

The movement of per capita income is given in Figure 2.2b. The per capita income of the Malthusian system *pci* quickly exceeds that of theRicardian system

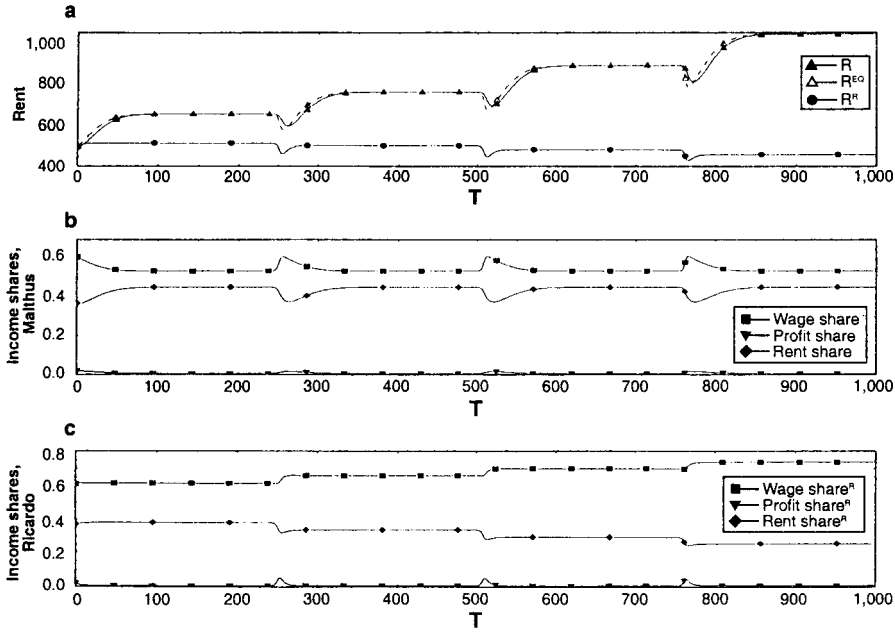

Figure 2.3 Movement of the principal macroeconomic variables of Ricardian and Malthusian systems over time: rent and income shares

pci^R. Per capita income is always half of the real wage in the Ricardian system. The movement of per capita income over time is more complicated in the Malthusian system because variations in the proportion of the population employed depend upon demographic behaviour as well as saving and investment behaviour. Effective demand determines the movement of the real wage and employment across time, which together determine the total income available to the labouring population. Per capita income, or the standard of living among the labouring class, is determined by changes in its total income, and the rate of population growth. The greater per capita income of the Malthusian system is in part due to the more restrained growth of population in that system, as shown in Figure 2.2c. The principal result of the endogenous growth of capital, labour and technology as far as Malthus was concerned was that the growth of output and population alternately outstrip each other so that the real wage alternately rises and falls along a secularly rising trend. This pattern is clearly visible in Figure 2.2b, which shows per capita income rising cyclically and secularly. A rising standard of wretchedness ensures that productivity shocks do not result solely in rising employment and population, but also in long-term improvements in the standard of living.

Malthus' opposition to the Ricardian doctrine that the interests of the landlord are opposed to those of the rest of society is clearly supported by the working of

this process over time. Figure 2.3a indicates that rent in the Malthus system is increased by agricultural technical change, while in the Ricardian system it has a rent-reducing impact. If a is increased in the technology shocks, corn rent will increase over time in both systems. That case has not been depicted here, as Ricardo did not include that concession in his chapter on rent, though he did do so in the concluding chapter on Malthus' rent theory (Ricardo 1951 – 73:412). The movement of shares in output is given in Figure 2.3b and c. Economic growth and a rising standard of living accompanies a growing rent share in the Malthusian model and a falling share in the Ricardian model.

POLICY IN THE INTERESTS OF SOCIETY

The compatibility of the landlords' interest with an improving standard of living became an important element of Malthus' support for agricultural protection. Another element was his belief that future events wouldeventually throw it back on its own resources.[15] The specification of technical change used here has the property that investment encounters diminishing returns once it absorbs or realises the improvement in technology generated in any particular wave; but across a longer period of time investment need never encounter diminishing returns, as long as it does not outrun the rate at which technology improves. This aspect of the specification reflects Malthus' belief that it would be difficult to suddenly and dramatically increase agricultural production without a similarly sudden and dramatic increase in price should the country be thrown upon its own resources, while moderate agricultural protection would suffice to maintain the rate of investment and technical change across time and allow agriculture to respond to greater demand without dramatic increases in costs of production, indeed with declining costs over time. The impossibility of rapidly catching up on forgone technological opportunities, together with a recognition of a commonality of interests between landlords and the mass of society, constituted the basis of his advocacy of agricultural protection.

Notes

1 Although Link's book was published in 1959, its content is advertised as substantially the same as his Ph.D. thesis completed in 1951 – 2.

2 Whether it is appropriate to follow Keynes in his assessment of the links between his theory and Malthus is a large subject beyond present capacities. The reader is referred to the remainder of Harcourt and O'Shaughnessy (1985) where some light is shed on the issue by examining Joan Robinson, Piero Sraffa and Richard Khan's contributions.

3 'The laws which regulate the rate of profits and the progress of capital, bear a very striking and singular resemblance to the laws which regulate the rate of wages and the progress of population' (1989a:370). See also Eltis 1984:143.

4 This approach ignores the imperfect competition determinants of investment examined by Harcourt and Kenyon (1976) amongst others.

5 See Harcourt (1969) for difficulties adhering to the definition and measurement of the profit rate outside of a one good model.

6 These terms may be the work of Malthus' anonymous editor.

7 As Costabile and Rowthorn (1984, 1985) have stressed, the real wage in this type of model is determined in the output market, not the labour market: a result consonant with Malthus' belief that eliminating nominal wage stickiness will not revive profits and growth if saving propensities remain too high. Link (1959: 67 – 8) has found only one passage which suggests that a reduction of the nominal wage might be of enduring benefit. That nominal wage movements will not bring recovery from stagnation is evident in the following passage:

> If, after labour has adjusted itself to the new level of prices, the permanent distribution of the produce and the permanent tastes and habits of the people should not be favourable to an adequate degree of [effectual] consumption, the clearest principles of political economy shew that the profits of stock might be lower for any length of time than the state of the land rendered necessary; and that the check to [rate of] production might
>
> be as permanent as the faulty distribution of the produce and the unfavourable tastes and habits which had occasioned it.
>
> (Malthus 1989a:446 – 7; 1989b:387 in brackets).

8 $d\pi/ds_1<0$, $d\pi/ds_2<0$

9 See Corry (1958; 1962:126) and Hollander (1962) for this position and Black (1967) for extensive references to the discussion.

10 On the Keynes-Malthus link see in addition Harcourt (1986:156 – 7).

11 Niehans (1963, 1975) presented an early analysis of growth with endogenous capital and labour supplies which has been unfortunately overlooked in this literature. Hicks and Hollander (1977), Goodwin (1979), Eltis (1980), Costabile and Rowthorn (1984, 1985) and Von Tunzelmann (1986) have presented formulations of endogenous population growth and capital accumulation in Malthus or Ricardo, some of which owe something to Wicksell's interest in population dynamics.

12 In Von Tunzelmann's (1986) model of Malthus, bounded rationality is insufficient to generate improvements in the long-run standard of living which are contingent upon the introduction of a rational expectations mechanism. Equations (14) and (15), however, indicate that bounded rationality can generate a rising trend.

13 Malthus' description of the evolution of the habits and customs which determine the standard of wretchedness seems to place more emphasis upon the past consumption experience of the lower classes rather than on social capillary action. For a very clear use of this formulation by a Ricardian, see McCulloch's (1975:99 – 100) treatment of taxes. Hirsch's notion of positional goods is usefully read against Malthus' standard of wretchedness on this point. Despite their differences, both generate the result that a rising (material) standard of living over time may confer

little psychic improvement. See Harcourt (1986:273 – 80) for a concise review of Hirsch's use of the concept.

14 Minami (1961) treats the standard of wretchedness as a constant, but even so presents a graphic examination of the potential for logistic waves of technical progress to secularly increase the standard of living.

15 Malthus was exceptionally creative in discovering various possible growth patterns in the countries which produce primary products which would cause the quantity of their exports to dramatically decrease and its price to dramatically increase. These possibilities may provide interesting scenarios for so called North-South growth models, but it is sufficient for our purposes here to accept his argument that a future war might throw Britain back upon her own supplies. This was an argument his contemporaries, who could well remember Napoleon's continental system, might have found persuasive, however wrong-headed it appears in the light of the ensuing Pax Britannica.

References

Amadeo, E.J. (1989) *Keynes' Principle of Effective Demand,* Aldershot: Edward Elgar.

Berdell, J.F. (1991) 'Essays on the Classical theory of advanced economic growth and trade', Ph.D. thesis, Cambridge University.

——(1995) 'The political economy of endogenous growth: a Malthus-like model of endogenous technological change and factor accumulation', Department of Economics Working Paper Series, De Paul University, March.

Black, R.D.C. (1967) 'Parson Malthus, the general and the captain', *Economic Journal,* 77:59 – 74.

Boserup, E. (1986) 'Shifts in the determinants of fertility in the developing world: environmental, technical economic and cultural factors', in D.Coleman and R. Schofield (eds) *The State Of Population Theory: Forward From Malthus,* Oxford: Basil Blackwell: 239 – 55.

Costabile, L. and Rowthorn, R. (1984) 'Malthus' theory of wages and growth', University of Cambridge, Faculty of Economics and Politics, Research Paper Series, no. 31.

——(1985) 'Malthus' theory of wages and growth', *Economic Journal,* 95:418 – 37.

Corry, B.A. (1958) 'The theory of the economic effects of government expenditure in English Classical political economy', *Economica,* new series, 25:34 – 48.

——(1962) *Money, Savings, and Investment in English Economics 1800 – 1850,* Basingstoke: Macmillan.

Duesenberry, J.S. (1949) *Income, Savings, and The Theory of Consumer Behaviour,* Cambridge MA: Harvard University Press.

Eltis, W.A. (1980) Malthus' Theory of Effective Demand and Growth, *Oxford Economic Papers,* 32:19 – 56.

——(1984) *The Classical Theory of Economic Growth,* Basingstoke: Macmillan.

Goodwin, R. (1979) 'Wicksell and the Malthusian catastrophe', in S.Strom and R. Thalberg (eds) *The Theoretical Contributions of Knut Wicksell,* Basingstoke: Macmillan: 62 – 70.

Harcourt, G.C. (1969) 'The accountant in a golden age', in R.H.Parker and G.C. Harcourt (eds) *Readings in the Concept and Measurement of Income,* Cambridge: Cambridge University Press: 310 – 29.

——(1982) *The Social Science Imperialists: Selected Essays,* London: Routledge Kegan Paul.
——(1986) *Controversies in Political Economy,* edited by O.F.Hamouda, Brighton: Wheatsheaf.
——(1988) 'Nicholas Kaldor, 12 May 1908 – 30 September 1986', *Economica,* 55: 159 – 70.
Harcourt, G.C. and Kenyon, P. (1976) 'Pricing and the investment decision', *Kyklos,* 29: 449 – 77.
Harcourt, G.C. and O'Shaughnessy, T.J. (1985) 'Keynes' unemployment equilibrium: some insights from Joan Robinson, Piero Sraffa and Richard Kahn', in G. C.Harcourt (ed.) *Keynes and His Contemporaries,* Basingstoke: Macmillan, 3 – 41.
Hicks, J. and Hollander, S. (1977) 'Mr Ricardo and the moderns', *Quarterly Journal of Economics,* 91:351 – 69.
Hollander, S. (1962) 'Malthus and Keynes: a note', *Economic Journal,* 72:355 – 9.
——(1986) 'On Malthus' population principle and social reform', *History of Political Economy,* 18:187 – 235.
Kaldor, N. (1955 – 6) 'Alternative theories of distribution', *Review of Economic Studies,* 23:83 – 100.
——(1980) *Essays on Economic Stability and Growth,* London: Duckworth.
Link, R.G. (1959) *English Theories of Economic Fluctuations, 1815 – 48,* New York: Columbia University Press.
Malthus, T.R. (1986) *The Works of Thomas Robert Malthus;* 8 vols, edited by E.A. Wrigley and D.Souden, London: W.Pickering.
——(1989a) *Principles of Political Economy:* Variorum edition, edited by J.Pullen, Cambridge: Cambridge University Press.
——(1989b) 'The alterations to the first edition (1820) of T.R.Malthus, *Principles of Political Economy,* in Malthus' manuscript revisions and in the second edition (1836)', with editorial comments, edited by J.Pullen, Cambridge: Cambridge University Press.
McCulloch, J.R. (1975) *A Treatise on the Principles and Practical Influence of Taxation and the Funding System,* edited by D.P.O'Brien, Edinburgh: Scottish Academic Press.
Minami, R. (1961) 'An analysis of Malthus' population theory', *Journal of Economic Behaviour,* 1:53 – 63.
Mishra, G.P. (1978) 'Malthusian economics of growth and development: a critique', *Indian Economic Journal,* 26:138 – 56.
Niehans, J. (1963) 'Economic growth with two endogenous factors', *Quarterly Journal of Economics,* 77:349 – 71.
——(1975) 'Economic growth and decline with exhaustible resources', *De Economist,* 123:1 – 22.
O'Leary, J.J. (1942) 'Malthus and Keynes', *Journal of Political Economy,* 50:901 – 19.
Pancoast, O. (1943) 'Malthus versus Ricardo: the effects of distribution on production', *Political Science Quarterly,* 58:47 – 66.
Rashid, S. (1977) 'Malthus' model of general gluts', *History of Political Economy,* 9: 366 – 83.
Ricardo, D. (1951 – 73) *The Collected Works of David Ricardo,* 11 vols, edited by P. Sraffa with the assistance of M.Dobb, Cambridge: Cambridge University Press.

——(1981) 'Malthus' principles and British economic thought', *History of Political Economy,* 13:51 – 79.

Sowell, T. (1963) 'The general glut controversy reconsidered', *Oxford Economic Papers,* 15:193 – 203.

Spengler, J.J. (1942) *French Predecessors of Malthus,* Durham NC: Duke University Press.

——(1965) 'Today's circumstances and yesterday's theories: Malthus on "services"', *Kyklos,* 18:601 – 13.

Vatter, H.G. (1959) 'The Malthusian model of income determination and its contemporary relevance', reprinted in J.C.Wood (ed.) (1986) *Thomas Robert Malthus: Critical Assessments,* vol. 3, London: Croom Helm, 115 – 20.

Von Tunzelmann, G.N. (1986) 'Malthus' "Total population system": a dynamic reinterpretation', in D.Coleman and R.Schofield (eds) *The State Of Population Theory: Forward From Malthus,* Oxford: Basil Blackwell, 65 – 95.

3
CASSEL ON CYCLICAL GROWTH

Mauro Boianovsky

It has long been acknowledged in the literature (see for example Robinson 1952: 42ff; Lundberg 1967:2 – 3; Phelps Brown 1981; Brems 1986:261ff; 1989:170ff; Niehans 1990:451ff; 1995:7, 19; Samuelson 1993:522) that Gustav Cassel put forward in his *Theory of Social Economy* (1918, tr. 1923 and again in 1932 from the fifth German edition; see Cassel 1932:61 – 3) the Harrod-Domar consistency conditions for steady growth—that is, g=s/v and g=n, so that n=s/v, where g is the rate of growth of income, s is the proportion of income saved, v is the capital-output ratio, and n is the rate of growth of population and labour force (see for example Hahn and Matthews 1964:783 – 4; Solow 1988:8 – 9). It has also been suggested in the literature that Cassel shared with the so-called Harrod-Domar model the assumption that the technology of the economy has fixed coefficients.[1] The notion that the Swedish economist assumed an exogenous capital-output ratio is behind the interpretations that either he had in mind an unstable economy with no mechanism to prevent increasing labour shortage or increasing unemployment (Brems 1986), or that his object was an economy on a full-employment growth path, which however would be reached only by a fluke or by 'accident' (see Tinbergen and Polack 1950:125).

I shall argue that the description of Cassel's as a 'knife-edge' growth model is not warranted by the textual evidence. The capital-output ratio is endogenous in Cassel, who consistently wrote the steady growth equation as v=s/n. The appearance of a fixed coefficient economy, which has misled the commentators, can be explained by the fact that Cassel (1932:61) assumed a 'uniformly progressive economy' in the passage where he presented his Harrod-Domar-like equation. In that case, the capital-output ratio is necessarily constant, since the economy is in balanced growth with unchanged relative prices and factor prices. One has to look in the chapters on the theory of interest and in his 1903 book in order to find Cassel's explanation of how the average capital intensity of the economy is determined through the influence of the price of the production factor 'waiting' on first the relative prices of goods produced with different capital intensities, and secondon the choice of the capital-labour ratio in the production of each good. Given s and n, the equilibrium rate of interest in the economy is such that v=s/n.

The pivotal role of the rate of interest in Cassel's discussion of uniform growth can also illuminate his interpretation of the business cycle as deviations—caused by the interaction between oscillations in the production of fixed capital goods and changes in the rate of interest—of the economy from its steady growth path. As has been pointed out by Ellis (1934:313 – 4), Hayek (1933:204 – 6), Haberler (1946:72ff) and Rose (1969:144 – 6), Cassel stressed, in his interpretation of trade cycles and crises, changes in the rate of interest instead of divergences between the 'market' and 'natural' rates typical of the Austrian approach. The rate of interest at the end of the boom is higher than the rate prevailing when the upswing started, which brings about losses for producers of new capital goods. According to Hayek and Haberler, under Cassel's assumption of a 'neutral' banking system, this can only be explained by an unexpected reduction in the average degree of saving at the end of the boom caused by changes in the distribution of income.

I submit below that, even if the degree of saving is constant during the business cycle, the upper turning point is explained in Cassel by the impact of an unexpectedly higher interest rate on the value of new capital goods already produced. While in his theory of economic equilibrium Cassel assumes, like Walras, that no capital goods are produced and sold at prices different from their production costs, in the study of the 'dynamics of the economy' (1932:534) he takes into account transactions in the market for capital goods at disequilibrium prices. The increase of the rate of interest during the boom brings the value of the new capital goods produced into equality with savings at prices below the incurred costs of production. The 'actual economic development', as Cassel (1932:533) put it, 'diverges from the previously assumed uniformity of progress'. This paper sets out to document and assess Cassel's discussion of what we may call a process of cyclical growth.

UNIFORM GROWTH AND THE CAPITAL-OUTPUT RATIO

One of the central contributions of Cassel's *Social Economy* is the extension—probably inspired by Marshall's (1990:306) suggestion[2]—of the notion of a stationary economy to the case of balanced growth (see Cassel 1932:32ff, 152ff), with constant relative prices and factor prices over time. Most of the book is dedicated to the study of equilibrium in stationary and uniformly progressing economies:

> The analysis of our problem must throughout be conducted in terms of this state of equilibrium, for the general theory of prices cannot be anything but a determination of the conditions necessary for the prices prevailing at any given moment to remain unchanged. This state of equilibrium is conceivable in either a stationary or a uniformly progressing society. The

> former case is of course by far the simpler; but the latter, too, must be taken into consideration, if we wish to be clear with regard to price-fixing in a society in which wealth is increasing.
>
> (Cassel 1932:94)

Cassel suggested three successive stages 'in the procedure from the abstract to the concrete': pure *static economy, uniformly progressive economy* (the 'quasi-static stage', when we can introduce 'such dynamic conditions as we are able to treat in a static form') and *economic dynamics* properly speaking (1925:27). Assuming away a continuous improvement in production methods, steady progress is only conceivable if population is growing at a certain rate:

> If we wish to study the simplest possible form of economic progress, we must select the case in which progress continues in the same way and with a constant amount of individual labour, and is thus merely increased in the same proportion as the population grows. If, in addition, we assume a uniform growth of population, we have the simplest case of a uniformly progressing economic system. If the annual percentage increase of population is given, then we know the rate at which the economic system as a whole is progressing, and we can speak of the definite rate of progress which is characteristic of that particular economic system…. In order that production may expand at the annual rate of progress thus obtained, it is necessary that the real capital of the system shall grow in the same ratio.
>
> (Cassel 1932:34)

Cassel was aware that the assumption of a constant income per capita was only a simplification, since one of the 'stylised facts'—as we may say after Kaldor (1989:230 – 1)—of economic growth is increasing income per capita. He suggested in 1903 an annual rate of growth of 2.5 per cent in aggregate production, of which 1 per cent was explained by the increase in population and 1.5 per cent by productivity growth (178). In the *Social Economy* Cassel adopted 3 per cent as the steady rate of output growth[3] and repeated the 1 per cent rate for population (Cassel 1932:62 – 3). As he later explained, 'in our typical progressive economy the product per unit oflabour shows an annual growth of something like 2 per cent, whereas the product per unit of capital remains constant' (Cassel 1935:138).

In the uniformly progressive economy the rate of interest is of course constant —he suggested a value between 3.5 per cent and 4 per cent in the *Social Economy,* and 5 per cent in the 1935 volume; the relation between the rates of interest and output growth in the steady state is determined by the degree of savings from interest income, as we shall see below—as Cassel used to point out against the then-prevailing view that the rate of interest tends to fall. With steady capital-output ratio and interest rate, the share of interest (or 'profits' from capital) in

income is also constant. The 'stylised facts' mentioned above—that is, steady growth of aggregate output, productivity of labour and capital-labour ratio; steady capital-output ratio; steady rate of interest; and constant share of interest in income—are perfectly consistent with Kaldor 1989. It should be noted that technical progress must be 'neutral' in the aggregate in order to be consistent with steady growth. Cassel assumed that the higher rate of growth of capital in some sectors is compensated by a lower one in others (1935:143, see also 28; and Harrod 1948:83, for a similar 'compensation' assumption).[4]

Cassel's equation for steady growth was introduced in his 1918 book as part of a discussion on 'the meaning of the continuous formation of capital' (1932: 61). He had criticised earlier in the book the 'classical' wage-fund view that the saving process consists in the accumulation of consumption goods that will be bought by the productive workers. As he put it, 'those finished goods which, in a state of continuous economic equilibrium, are not consumed because the extent to which wants may be satisfied has voluntarily been limited, are never actually produced' (1932:37). The production factors that would be used for the production of consumption goods are instead transferred to the production of real capital. 'Saving…consists in a particular method of applying factors of production. There is…no concrete object which is saved in the sense that its consumption is deferred' (1932:36 – 7).[5]

The meaning of saving as the part of income used to purchase the newly produced real capital is particularly clear in the 'uniformly progressive economy', as Cassel set out to show (1932:61 – 3). He assumes a 'money economy'—in the sense that money is used as a 'common scale of reckoning'—where 'abstract total capital' C at a moment of time means, as Robinson (1952: 43) put it, 'all the goods in existence at that moment, valued at their prices in terms of a unit of final output', Cassel further assumes that in the uniformly progressive economy capital increases at a steady annual rate of p per cent. If we bear in mind his definition of uniform growth, p must be the rate of population growth or, more generally, the rate of growth of the effective labour force, augmented by the rate of neutral technical progress. Cassel is not explicit about that in pages 61 – 3, but his numerical illustration and his reference to his previous discussion of steady growth clearly suggest that p is given. He uses I and 1/s as annual income and 'degree of saving' respectively,which is assumed to be constant. He then writes the equation I/s=(p/100)C, expressing the equality between saving and investment. Alternatively, the equation is written as an expression for the level of income for a certain capital stock at a moment of time, that is, I=(sp/100)C (1932:61).

What conclusions did Cassel draw from these equations? As he put it in the Swedish edition, 'with both p and s constant, this means that income stands in an invariable ratio to capital' (1934:86).[6] It is clear enough that a steady capital-output ratio is not an assumption, but a corollary of the model of uniform growth:

> We therefore come to the conclusion that, in the uniformly progressive exchange economy, *the total income as well as both its parts—consumption and capital accumulation—increases in the same percentage as the capital*.
>
> (Cassel 1932:62; original emphasis)

Using modern notation (see above, where Cassel's *p*/100, 1/*s, I* and *C* are respectively *n, s, Y* and *K,* with *v* defined as *K/Y*), Cassel's first equation reads *sY=nK*. In this equation, *n* and *s* are the parameters, while *Y* and *K* grow continuously along the steady growth path. Therefore the equation can either be used to find the income level at any moment for a given initial capital stock,[7] which he illustrated on the middle of page 62; or it can be used in the form *v=s/n,* which Cassel did at the bottom of the same page:

> If we assume, for example, that the rate of progress is equal to 3 per cent, and that therefore the capital C increases annually by 0.03C, and if we further assume that one-fifth of the annual income is saved, then, according to what we have said, the income equals 15 per cent of the capital.

Brems (1989:171) is aware that Cassel writes v=s/n, instead of n=s/v, but insists that it is only a 'rearrangement'. He therefore suggests that Cassel assumes values for v and s in order to find the rate of growth, which is the opposite of Cassel's reasoning. (Besides, Brems wrongly dissociates the 'rate of progress' from the rate of population growth.)

Further evidence that Cassel did not assume fixed coefficients can be found in his reaction to Paul Douglas' 1933 pioneer econometric estimation of what would become known as the Cobb-Douglas production function. Cassel (1935: 134) reproduces Douglas production function as $p=x^u y^v$, where p is 'the total production of a self-contained community', x and y represent labour and capital respectively, and $u+v=1$. He argues that, since in a uniformly progressing community capital and income grow at the same rate, we must have p=cy, where c is a constant. Douglas's equation is then written as $cy=x^u y^v$, which takes the form $c=(x/y)^u$. It is clear enough thatfor Cassel the steadiness of the capital-output ratio (and in this case of the capital-labour ratio as well) is a property of the steady state, not of the production function.

One of the well known corollaries of full-employment growth models, of the kind discussed by Cassel and Solow-Swan, is that the rate of growth is determined by the rate of growth of the 'effective' labour force. The saving propensity can only affect, in the steady state, the equilibrium capital-output ratio and the labour-output ratio (that is, income per head), not the rate of growth. Cassel, however, often suggested that the 'rate of progress' is determined by the 'degree of saving' (see for example 1932:250). He discussed in 1934 (88) the consequences of a reduction of the degree of saving from 1/5 to 1/10 in an

economy which is growing at 3 per cent, so that the initial steady output-capital ratio is 15 per cent, as in the illustration quoted above. According to the pure logic of the model, the only effect should be an increase of the output-capital ratio to 30 per cent (brought about by an increase of the steady state rate of interest) without any impact on the rate of growth. Instead, Cassel suggests that the rate of growth would be reduced to 2 per cent, and the output-capital ratio consequently increased to 20 per cent. Cassel's attempt to ascribe a positive, determinant role for saving in the growing process can be seen in part as ideologically motivated, but it can be explained in terms of the model by introducing the notion that the rate of growth of labour is not independent of the rate of accumulation, as Swan (1956:338 – 9) suggested. This would happen if the supply of labour were elastic in relation to the level of output per head, reflecting a Malthusian mechanism, or a potential supply of migrant labour, or even disguised unemployment. Since Cassel contended, in his chapters on economic fluctuations, that historically the migration of rural workers accounts for the supply of labour during periods of intense growth (Cassel's 'industrial reserve army'; 1932:564), the mechanism pointed out by Swan can be readily incorporated into his framework. To the extent that the labour supply is elastic in the neighbourhood of a certain income per capita, an increase of the saving ratio will raise the rate of growth and have proportionally less impact on the capital-output ratio, as we can see in Cassel's 1934 exercise. Behind all this, however, is the—until now—tacit assumption that the rate of interest brings the capital-output ratio to its equilibrium level, as we shall see next.

'WAITING' AND THE RATE OF INTEREST

Cassel (1935:24ff) used his notion of uniform progress to criticise Böhm-Bawerk's view that the 'average period of production' tends to rise through time. He suggested that the capital-output ratio is a measure of capital-intensity in the economy in equilibrium:

> We ought to observe that a more natural and reliable measure of the importance of capital in the social economy could be obtained simply by calculating the quotient between the capital and the income of the economy. If we express both capital and income in terms of money, and *if we assume a state of equilibrium in which all prices are fixed and where therefore real capital has a price corresponding to its cost of production,* this quotient has a definite quantitative meaning, and is obviously an important characteristic of the social economy.
>
> (Cassel 1935:22; also 1934:89 – 90; emphasis added)

Cassel's suggestion that the average capital-output ratio provides a measure of the degree of roundaboutness of the economy can be better appreciated if we bear in

mind his concept of 'waiting' as a primary production factor. Capital goods are produced by labour, but, because of their durability, they can only be used if the owner waits for their services. Even though Cassel acknowledged that the production process itself is time-consuming (which he associated with 'circulating capital'), he usually stressed that the amount of waiting required on the grounds that the consumption of durable goods takes time is 'by far' greater than the amount demanded because production takes time (1934:92; 1932:217 – 8; as Wicksell [1919:240] noticed, this is the case because Cassel assumed a growing population accompanied by a rising demand for fixed capital). 'Waiting' is defined as 'the foregoing, for a certain time, of the consumption of an existing capital' (1932:199; see also 1903:133 – 4).[8] This definition was adopted by Dorfman (1959:359), who, however, did not refer to Cassel (cf. Blaug 1985:516 – 7). As Dorfman points out, 'waiting' should be regarded—together with labour—as a primary (non-produced) production factor, since the owner of a durable capital good has the alternative of selling the fixed capital good at the beginning of a time interval during which the machine (actually, it can also be a durable consumption good, such as a house) is used and buy consumption goods. Hence by keeping the capital good, the owner performs an amount of waiting measured by the current value of the capital good (in terms of consumption goods) times the length of the time spell.

> Waiting, being synonymous with use of capital, is a quantity of two dimensions, measured by the product of a certain sum of money multiplied by a certain time, the unit for the measurement being the waiting for one pound over one year. Thus a company working with a capital of one million pounds uses yearly a quantity of waiting of one million units.
>
> (Cassel 1903:90)

Dorfman used this concept to show that the average period of production for economies with two sectors (consumption goods and durable capital goods) can be measured by the average capital-output ratio when the periods of investment of labour and of waiting are taken into account (1959:218). In the same vein, Cassel (1903:90 – 1; 1932:218) had suggested that the quotient between the total capital employed and the total amount of yearly expenditure could be used as a proxy for the waiting-labour ratio in different sectors. His empirical findings confirmed his view that the proportion between waiting and labour varies widely among different goods, and that consequently, a change of the price of waiting (the rate of interest) would affect relative prices in the economy.

As we saw above, Cassel was aware that the capital-output ratio is a reliable measure of capital-intensity only if the economy is in equilibrium—whether stationary or steady—with capital goods valued at their costs of production. This will happen if the producer of a new capital good is able to get enough waiting to

cover its production costs—in other words, if he can sell the machine for a price corresponding to its cost:

> As soon as it is intended to make a durable article, the capital-disposal [a synonym for 'waiting'] that is necessary if the waiting for its services is to be done must be regarded in the same way as anything else required for its production…. If, on the other hand, the durable article is finished, then one must wait, and the necessary capital-disposal may or not be found. If no one is willing to pay more than half the cost of the article, there is a loss which must be borne by the producer. The buyer has then, of course, only half the capital-disposal to offer. But then the conditions necessary for the continuance of production are disturbed.
>
> (1932:200; see also 94 – 5)

Cassel's emphasis on the study of the conditions necessary for the maintenance of equilibrium—including the condition that the number of equations is equal to the number of variables in the general equilibrium system—often led him to neglect, most notably in his *Social Economy,* the investigation of how the market solves the equations, namely what Walras called the tâtonnement process (cf. Patinkin 1965: note B, esp. 536, n18). The task here is to show that, if the system is initially in equilibrium, 'any arbitrary departure from this position generates dynamic market forces which bring it back' (*ibid.:* 539), that is, a problem of stability. This is particularly relevant if Cassel is to be able to show not only that aggregate uniform growth is possible when the capital-output ratio is endogenous, but also that the market brings that ratio to its equilibrium level. It is necessary, in other words, to show how, starting from a position of steady growth equilibrium, an arbitrary change of the rate of interest will generate an excessdemand (or supply) for the production factor waiting, which will bring the rate of interest back to equilibrium.

The relation between the demand for waiting and the rate of interest is explained by Cassel according to what he used to call the principles of 'scarcity' and 'substitution' (see 1903: ch. 3; 1932:216ff). The former consists in the effect of the rate of interest on the relative prices of goods produced with different capital (or, better, waiting) intensities. It is the only principle he incorporated into his formal presentation of general equilibrium, where he assumed fixed coefficients in the production of each good.[9] A reduction of the interest rate brings about an excess aggregate demand for waiting, as the demand for capital-intensive goods rises. The 'principle of substitution', on the other hand, has to do with the influence of the rate of interest on the choice of the production method and on the durability of capital goods:

> A local board might be in doubt whether to build a bridge cheaply, to last for fifteen years only, or at double the expense, to last for sixty years. We

> might, for the sake of simplicity, assume the cost of repairs to be the same in either case. Then it is clear that, in the case of the more expensive bridge, a larger sum has to be paid on account of waiting, and a less on account of depreciation. Hence a certain quantity of waiting might be said to be substituted for other factors included in 'bridge-building'. The kind of bridge to be built will, again, depend wholly on the rate of interest. The calculation is simplest if we suppose the board in either case to cover the cost by a loan repayable by an annuity for a fixed period…the calculation shows that the annuities are nearly equal when the rate of interest is 4.125 per cent. With a higher rate it will prove more advantageous to build the bridge on the cheaper plan; but if the rate be lower than 4.125 per cent the more expensive bridge will be preferred.
>
> (Cassel 1903:111 – 12)

It should be noted that the rate of 4.125 per cent in this illustration corresponds exactly to Irving Fisher's (1907:153 – 4) concept of 'rate of return on sacrifice' (or 'rate of return over cost' in the 1930 version), that is, the discount rate at which the present net values of two alternative income-streams are equalised. In his writings on the business cycle Cassel (1932: 639 – 40) usually adopted a rule closer to Keynes's concept of 'marginal efficiency of capital' for a single investment option (cf. Blaug 1985:529), that is, the rate of discount that equalises the replacement cost of the machine to the present value of the expected receipts.

The supply of waiting (aggregate saving), according to Cassel (1903: ch. 4; 1932:232ff) is not a function of the rate of interest, unless interest falls below a certain critical level. Cassel, in contrast with Fisher and Wicksell[10] did not use utility maximisation in his discussion of savings. This is consistent with his general criticism of marginal utility in the theory of demand (see Samuelson 1993:517, on the limitations of what Samuelson calls 'Cassel's revealed-demand approach'). He assumed that the main reason for saving is to provide for the individual's future needs. In the case of a 'capitalist' who plans to live in the future on an annuity, a reduction of the current interest rate increases the length of time he will have to save in order to obtain the annuity. Given his life expectancy, if the rate of interest falls below a certain level (between 2 and 2.5 per cent, according to Cassel 1932; see the table at 243, where he considers different lengths of life for individuals of various ages) the representative individual will start to consume his accumulated capital (see Spengler 1969:162 – 3). Therefore, this critical level decides the rate of interest in the stationary state. Generally speaking, Cassel assumes that the 'degree of saving' s is constant, apart from changes in the very long term as 'civilisation advances'.

Cassel (1928, 513 – 14; cf. 1903:79 – 80 and 1925:100 – 1) used his principles of scarcity and substitution to argue that, starting from an equilibrium position, 'if the rate of interest were lower than it actually is, forces would be called forth compelling the rate of interest to rise again to the actual height required for the

maintenance of equilibrium'. A reduction of the rate of interest would cause an increase in the demand for durable goods, with an ensuing excess demand for waiting, which 'would immediately force the rate of interest to rise again'. The crucial—but only implicit—assumption here is that in this tâtonnement actual transactions are made only at the equilibrium level of the rate of interest, where prices of capital goods are equal to their production costs.

Furthermore, it is worth noting that Cassel did not put forward (with an exception to be mentioned below) a mechanism akin to the Solow-Swan capital-deepening process. In their formulation, if initially the capital-output ratio v is below the level required to make $n=s/v$, this means that the capital stock is growing faster than population and the rate of interest is falling. The rise in the capital-labour ratio (and so in the capital-output ratio, if the production function is 'well behaved') continues until it has increased the 'amount of widening investment that is required in each period to keep pace with the increase in population' to a point where full-employment savings are fully absorbed by widening investment (Hahn and Matthews 1964:809 – 10). The Solow-Swan process consists in a succession of stationary states with different amounts of capital goods along an equilibrium path where expectations are always confirmed. Cassel, on the other hand, did not conceive of convergence from an initial disequilibrium position as a succession of stationary states—where relative prices would be changing over time, in contrast with his definition of equilibrium quoted above—but investigated instead the mechanism that would prompt the system back to the initial equilibrium situation after departing arbitrarily from it. Nevertheless, the outcome is the same in both cases. It is possible,however, to find a single passage in Cassel (1903:121) where he discusses a process of falling interest rate, not unlike Solow-Swan , caused by a 'large and continuous growth of capital' in excess of the demand warranted by population growth (capital-widening). The ensuing increase in the capital-output ratio would make investment demand equal to the flow of net saving, and therefore interrupt the fall: 'Every fall of this rate [of interest] will widen the field for the use of durable instruments and thus call out forces counteracting the fall' (1903:121).

According to the principles of 'scarcity' and 'substitution', the demand for waiting is a negative function of the rate of interest.[11] Population growth and technical progress shift the curve over time. At the same time the supply of waiting—which is not a function of the rate of interest, except below the critical level decided by the average length of life—increases according to the growth of full-employment income. The 'chief determining factors' of the rate of interest are, accordingly, the actual length of human life, the rate of increase of population, and technical progress (1932:254). Cassel consistently related the level of the rate of interest to economic growth. It was in his first book, while examining the determinants of interest in a socialist economy (see Cassel, 1900: 124ff; repeated in abridged form in 1903: ch. 6), that Cassel first put forward the

notion that 'the determinant factor of the rate of interest is essentially the relative progress of the economy' (*ibid.:* 132).[12] That relation is straightforward in a socialist economy, where, by definition, the whole income from interest is saved and invested by the state.

> Let us suppose the community to increase its capital every year 1 per cent on account of the growth of population, and 1.5 per cent on account of real progress. This is probably not more that what is very often done in our present society. This 2.5 per cent represents a certain quantity of labour which can be paid only by raising a corresponding sum in the form of interest on the capital of the society. Hence it follows that the Socialist community under the given circumstances would have to charge interest at 2.5 per cent; this interest being counted as a part of the cost of production, and therefore of the selling price of every commodity.
>
> (1903:178; cf. 1900:128)

Cassel's proposition that the rate of interest is equal to the rate of growth when interest is all saved and workers consume the whole of their income is of course reminiscent of Von Neumann's seminal contribution and also of the famous 'Golden rule'. It should be noted, however, that, in contrast with Von Neumann, the rate of growth in Cassel is exogenous (see Kurz and Salvadori 1993:136), and that he did not associate that result with maximal consumption per head (nor did Von Neumann, for that matter). The rate of interest will not generally be equal to the rate of growth in capitalisteconomies, as 'we are not bound to that peculiar principle characteristic of the Socialist community, that the total interest income of a year must be equal to the increase of capital in that year' (Cassel 1903:179).[13] Assuming, as Cassel and his contemporaries usually did, that workers do not save, the relation between the rates of interest and steady growth can be easily seen if we use Cassel's equation $nK=sY$ discussed above. In this case, we can write $nK=s_pP$, where P is income from interest (or 'profits') and s_p stands for the proportion of profits saved. Hence $P/K=n/s_p$, with the 'socialist' result that the rate of interest P/K is equal to n if s_p is 1. Cassel used (1935:138 – 9) an example where n is 3 per cent, s is 15 per cent (so that $K/Y=5$) and assumed furthermore an interest rate of 5 per cent, which implies that P/Y is 25 per cent. The proportion saved out of profits is, therefore, 60 per cent. In his *Social Economy* n is also 3 per cent, but he mentioned (1932:255) 4 per cent as the 'usual' rate of interest, which implies that 75 per cent of profits are saved and that the share of profits is about 26.7 per cent of aggregate income. Brems' (1986:162 – 3) suggestion that Cassel never related the equilibrium values of the rates of interest and growth is not warranted by the textual evidence. It should be noted that Cassel (1935) used the numerical exercise to illustrate how, against Douglas, the variables are determined in steady state irrespective of marginal productivities (see Pasinetti 1974: 125 – 6, for a similar argument).

DISEQUILIBRIUM AND THE BUSINESS CYCLE

Cassel used the concept of 'uniformly progressive economy' for two distinct purposes. The first one, which we have been discussing so far, is the study of properties of steady state economies, including comparisons between different steady states when one of the parameters changes—an exercise in 'comparative dynamics', to use the term suggested by Hahn and Matthews (1964:781) in the same context. Alternatively, Cassel referred to equilibrium growth as an 'average' rate of growth around which oscillates the actual rate (e.g. 1935:95). This is the concept relevant to the study of business cycles, which, as Cassel often pointed out, belongs to 'the dynamics of the economy' (see Hicks 1985:9 – 10, for the distinction between the two uses of the notion of steady growth).

One of Cassel's favourite exercises in 'comparative dynamics' is to examine the effect of a change in the rate of population growth on the rate of interest (see 1903:55; 1932:221). A reduction of the rate of population growth brings about a decline of the rate of interest and an increase of the capital-output ratio to its new steady state level, in such a way that the equality n=s/v remains. It is instructive to compare this with a similar exercise in economic growth carried out by Keynes in his 1937 lecture on 'Some economic consequences of a declining population'. Keynesinvestigates what would happen if, in an economic system initially in growth equilibrium with a capital-output ratio equal to 4 and 15 per cent as the proportion saved from full-employment income, the rate of population growth falls significantly (he actually considered the case of a stationary population, but this does not affect the argument). Given those values —deduced by Keynes from data for the period 1860 – 1913, the same period studied by Cassel—the capital stock is increasing at an annual rate of 4 per cent, which means that the demand for capital must be also grow at 4 per cent if full-employment equilibrium is to be preserved. According to Keynes (1937:126 – 30), the three sources of demand for capital are population, 'standard of life' (as determined by inventions that do not change capital per unit of output) and 'capital technique' (the capital-output ratio, determined by the rate of interest). Given the degree of saving and the capital-output ratio, a decline in the rate of population growth means that the natural rate of growth is less than the warranted rate of 4 per cent (Keynes did not use Harrod's terms 'natural' and 'warranted', but the concepts certainly correspond to the ones used by Harrod two years later), with ensuing excess saving and unemployment. Keynes, for familiar reasons, did not believe the rate of interest could be reduced below its average value during 1860 – 1913 in order to elicit the necessary increase of the capital-output ratio and compensate for the decline in population growth. Chronic unemployment, on the other hand, is no part of Cassel's theoretical framework, since the rate of interest is flexible. This flexibility, however, does not prevent the economy from experiencing cyclical unemployment caused by oscillations of the rate of interest around its equilibrium level, as we shall see below.

According to Cassel (1932:62), the 'principle' that income and capital grow at the same rate is 'approximately correct in normal circumstances for every economy. Only during transition periods will there be a material difference between the rate of increase of capital and of income' (1932:62). 'Transition periods' here do not mean a Solow-Swan equilibrium path of convergence of the capital-output ratio towards its equilibrium level, but deviations of the economy from growth equilibrium, accompanied by an imbalance between the flow of saving and the production of new capital goods. Those are not 'normal circumstances', since capital goods are not sold at prices corresponding to their costs of production, and expectations are not confirmed (cf. the distinction suggested by Hahn and Matthews 1964:782, between 'equilibrium' and 'disequilibrium dynamics'; the latter is not part of the Solow-Swan model).

Cassel's extensive discussion of trade cycles in Book IV of his *Social Economy* is in great part a development of a long article on economic crises published in Swedish in 1904, where he reacted against J.A.Hobson's under-consumption approach and made use of the concept of 'waiting' (for the services of durable goods) as a production factor, introduced a yearbefore in his *Nature and Necessity of Interest*. Cassel pointed out that the recognition of the possibility of a discrepancy between 'capital accumulation' (in the sense of 'creation of concrete capital', i.e. production of new durable capital goods) and the supply of saving (in the sense of 'waiting' for their services) could 'lead to a complete explanation of oscillations between good and bad times' (1904:61). It is clear enough that Cassel is here departing from the view—adopted in 1903 and in 1932, Books I and II—that the economy is always in equilibrium, with the rate of interest immediately adjusted to the level necessary to equilibrate the demand and supply of waiting. He now describes the business cycle as oscillations in the production of fixed capital goods, accompanied—but not prevented—by changes in the rate of interest as an expression of the price of the factor 'waiting'.

> The deepest cause [of economic crises] is the scarcity of a production factor, which is called waiting. That scarcity shows itself not only in the difficulty to raise loans, but also in the difficulty to find buyers for real capital goods. He who buys a house or pig iron is in effect performing a certain amount of waiting. But, if this factor is scarce, there is also a shortage of buyers, and prices of capital goods fall.[1]
>
> [1] The scarcity of waiting asserts itself in the fact that nobody will take over the existing real capital—not in that one does not believe that the capital goods are able to yield their estimated return, but just that one does not wish to wait.
>
> (1904:69)

The footnote added by Cassel indicates that the upper turning point is not caused by 'capital saturation'—in the sense of an over-production of capital goods in

excess of demand from final consumers, as suggested by Arthur Spiethoff and others (see Hansen 1951:315). He pointed out that

> the services of fixed capital, which are the object of the demand of the consumers, are generally not at all in excess even in the last part of the boom, [but], on the contrary, the fixed capital has to be used to the utmost to meet the demand.
>
> (Cassel 1932:649)

What happens is that people are not willing to forgo their current consumption in order to keep and use the newly produced durable goods. Since society as a whole cannot convert fixed capital goods into consumption goods by selling (cf. Dorfman 1959:361 – 2), there follows a fall of their prices below production costs. As Cassel (1932:649; cf. 1904:70) put it, economic crises result from

> *An over-estimate of the supply of capital, or of the amount of savings available for taking over the real capital produced.* What is really over-estimated is the capacity of the public to provide savings in sufficient quantity.
>
> (1932:649, original emphasis; cf. 1904:70)

He sometimes referred to the 'capital market' as 'a market on which the fixed capital produced is offered for sale' (1932:623), as if capital in the physical sense was the object of transaction (cf. Wicksell 1919:239), but he was aware that there is no such a market:

> It is not an easy matter to give a general idea of the fluctuations of the prices of fixed capital, for there is no proper market for the commodities in question. The greater part of fixed capital is, however, represented in paper values of one kind or another on the Stock Exchange. The demand for savings capital with which to purchase the fixed capital that *has been produced* is expressed partly as a demand for the means of buying shares, and partly as a demand for loans. The ownership of fixed capital is directly represented by shares, but the loans which reach the Stock Exchange in the form of bonds bearing a fixed interest also serve, from the economic point of view, as a means for taking over newly produced fixed capital.
>
> (1932:630 – 1; emphasis added)

The determination of the rate of interest in Cassel's business cycle can be better appreciated with the help of Lundberg (1937:216ff),[14] who also assumed a banking system that only finances current production and does not take over fixed capital. According to Lundberg (1937:218), the entrepreneur takes the rate of interest i_{t-1} prevailing at the end of period (t−1) as given when planing his

production of new durable goods during period t , that is, he 'expects this rate to be ruling when he has to issue the [bonds] after having built the [capital good]' (Lundberg actually uses 'houses' and 'mortgages' in the illustration). The costs incurred in producing the capital good during period t are, therefore, covered by the issue of bonds in period t+1. Cassel (1932:638) also refers to this process of funding shortterm loans. As Lundberg points out, banks are assumed to be 'neutral', in the sense that they do not affect the long-term rate of interest. The bonds issued are bought in the stock exchange with savings performed in period (t+1), while the demand for bonds is equal to the value of the investment done in period t. If they differ—as it will be the case if investment is growing—the rate of interest prevailing in (t+1) when bonds are issued will be higher than the rate in (t−1), when plans were made. This timing in the determination ofthe interest rate is clearly assumed by Cassel in his hypothesis about crises as result of an over-estimation of savings:

> We must bear in mind that this capacity [of the public to provide savings] has to be estimated many years in advance, since, on the average, there are several years between the time when the work is planned and the time it makes its full demand upon the community's savings. The individual entrepreneur has no other means of judging the condition of the capital market except the rate of interest. The rate, however, is low…during the depression and the first part of the trade boom, since the demands for capital-disposal which result from the increased activity of entrepreneurs in the production of fixed capital, do not yet make themselves fully felt.
>
> (1932:649 – 50)

The divergence between saving and investment and the ensuing change of the rate of interest are not caused by an interference of banks in the credit market—that is, by a Wicksellian difference between the market and natural rates of interest—but by the fact that 'the supply of savings during one period is used to take over investments made during the preceding period', as Lundberg (1937: 225) put it. This is fully consistent with Cassel's concept of saving as supply of 'waiting' demanded in connection with the use of durable fixed capital, that is, after the capital good has been produced. In his 'general theory of prices', as we have seen above, it was necessary to assume that interest rates are always at equilibrium, without real losses for producers of capital goods. This corresponds, of course, to Walras' (1954: 292 – 3) assumption that producers use 'tickets' instead of entering into actual transactions in disequilibrium, in contrast with the trade cycle.

> We must…investigate the changes which the capital market experiences during the various phases of the trade cycle. To get a clear idea of these changes, we will first assume that there is no change in the use of income,

> or, in other words, that the ratio of the income consumed and the income saved remains unchanged. In that case both the income consumed and the newly created capital increase at the same rate as total income. But during the trade boom the production of fixed capital increases at a faster rate than the total income. The prices of the material capital goods must obviously fall in the conditions supposed. This movement will continue until equilibrium is restored between the total value of the newly produced real capital and the income available for the purchase of it.
>
> (Cassel 1932:622 – 3)

The 'equilibrium' between the value of new capital goods produced and saving, however, is not actually an equilibrium, since it is brought about by a reduction of the prices of capital goods below their production costs. As Schumpeter (1954: 1019) has pointed out in another connection, the real equilibrium condition in both Walras' and Cassel's approach to the capital market is not the equality between saving and investment, but the equality between the amount saved and invested and the aggregate costs of the new capital goods. As Cassel suggested, if savings are not enough to buy the real capital produced,

> There must then be a sudden fall in the value of fixed capital, and entrepreneurs must find it extraordinarily difficult to get the capital they need, either by loan or selling securities…. When it becomes clear that they have been mistaken in this respect, there is bound to follow a widespread inability to meet the liabilities incurred…. There is bound to be a general crisis.
>
> (1932:650)[15]

The ensuing contraction in the production of capital goods brings about the depression period, marked by widespread excess capacity (see the detailed discussion of the empirical evidence at 588ff). Cassel (1904:74ff; 1932:596ff) pointed out that even a slight fall (or interruption of the growth) of the demand for consumer goods would provoke a strong decline in the production of durable goods used as inputs in the capital goods industry—an early formulation of the 'acceleration principle' (see Hansen 1951:312), which Cassel consistently associated with the durability of capital goods. The 'acceleration principle' is only one element in Cassel's framework though, not the deciding factor in the business cycle. The reduction in the demand for waiting during the depression brings down the rate of interest, which stimulates investment and leads to the starting of the boom period. There is thus, according to Cassel (1932:641) 'a reciprocal action between the rate of interest and the progress of the trade cycle'. This 'reciprocal action' produces damped fluctuations, which will be kept alive by the irregularity of the determinant factors of investment demand through time —that is, mainly technical progress, but also 'the exploitation of new countries'

(which changes the demand pattern towards capital-intensive goods in hitherto 'semi-civilised' lands) and population growth (1932:644 – 7).[16]

> The wavelike movement is essentially the effect of a continual struggle between the pressure for progress and the scarcity of the capital resources required for the constructive work in which progress must manifest itself. In this interplay of forces the rate of interest is an essential factor, and its fluctuations reflect the changesthat the capital market is always exposed to in a progressive economy.
>
> (Cassell935:101; cf. 1932:647)

He further pointed out that this wave movement alternately 'rises above and falls below the curve of the uniform progress', which represents an average development (1933:1). The rate of growth, therefore, 'must, in the upward swing of the cycle, be greater than normal' (1932:493). Cassel, however, did not spell out what are the determinants of the actual rate of growth as opposed to the 'normal' one. From his emphasis on changes in capacity utilisation over the trade cycle, it is possible to infer that fluctuations in the rate of growth are accompanied by changes in the actual capital-output ratio (as opposed to the equilibrium on when prices are equal to production costs). During the boom period, 'better use is made of the durable means of production' (1932:589), not only in the capital goods industry, but also in the consumption goods sector, since the revival of activity, 'which begins in the capital-producing industries, spreads also, though in a progressively declining degree, to other branches of industry' (*ibid.:* 617). The increase of the degree of capacity utilisation—operated by 'the new workers recruited from among the ranks of the unemployed, from young people hitherto not employed, or from the agricultural production', as he explains at 617—means a (unexpected) reduction of the actual capital-output ratio v. Hence, given the degree of saving s, the actual rate of growth s/v will be higher in the boom when compared to the 'average' rate at normal capacity utilisation (which, as we have seen above, is equal to the rate of growth of the effective labour force). And vice-versa in the depression period, when the actual capital-output ratio rises because of excess capacity, which brings down the rate of growth below its 'normal' figure (cf. Matthews 1959:238 – 9). We are here far from the tranquil world of the 'uniformly progressive economy'.[17]

FINAL REMARKS

The interpretation advanced here of Cassel's explanation of the upper turning point is not based—in contrast with Haberler (1946), Hayek (1933) and Rose (1969)—on the assumption that the degree of saving falls in the final part of the boom. Although Cassel does refer to that in connection with changes in income distribution during the cycle, he makes it clear that the fall in saving only

'accentuates' excess demand in the capital market, 'and the interest rates must rise still higher than was necessary in our first assumption' (1932:620 – 1, 624). In other words, the fall in s is a sufficient but not necessary condition or the crisis. Hayek's interpretation of Cassel is a by-product of the Austrian's fundamental assumption—before his so-called'transformation' in 1937—that prices carry all the information necessary to bring about a state of equilibrium, except if the money rate of interest is set at the 'wrong' level by the banking system, or if 'temporary changes' (not perceived as such by producers) in the equilibrium value of the interest rate—caused by a temporary rise in s in the initial stages of the boom—are introduced (cf. Hayek 1933:81 – 2, 204 – 5). Nevertheless, it is true that, even, though the textual evidence supports the alternative interpretation put forward above, Cassel never spelled out the Walrasian tâtonnement in the equilibrium model of the capital market, let alone the disequilibrium process characteristic of the business cycle.

One of the striking features of Cassel's process of cyclical growth is that money is conspicuous by its absence. As he pointed out, the 'money economy' assumed in his discussion of capital accumulation in the uniformly progressive economy (1932:51ff) is 'monetary' only in the sense that all values are expressed in money of account, since money was not assumed to be demanded for transaction purposes: 'As we then ignored the existence of media of payment, we came to the conclusion that every item of income that the community receives is, right from the time of its creation, invested in real capital' (1932:440).

When money is used as a store of value, the former equality is no longer valid, since savers are not restricted to the accumulation of real assets. Needless to say, Cassel did not develop an economic growth model for a monetary economy (this was done only in the 1960s by James Tobin and others; cf. Solow 1988: ch. 4) but the positive relation between expected falling prices and demand for money was an essential element of his interpretation of the deep depressions of the early 1920s and early 1930s. When prices begin to fall,

> the recipients of income will then adopt a policy of holding back their money. They reduce their outlay in the hope of having the opportunity later on of buying more cheaply, and they tend, for the time being, to keep their savings in money form.
>
> (1932:444)[18]

The 1929 crisis, according to Cassel, did not belong to the 'normal' course of the business cycle, but to the realm of pure monetary phenomena caused by monetary policy mistakes. From that point of view, it represented an absolute discontinuity in the process of cyclical growth (cf. 1934:87). He dismissed both the underconsumptionist and the excess investment ('the central feature in usual crises') explanations as inadequate, and stressed that 'the course of economic events in the United States is essentially a pure process of deflation, quite

distinct from ordinary economic movements' (article in the October 1932 *Quarterly Report:* 62; see also his *Social Economy:* 506 – 8; cf. Carlson's [1993: 166] suggestion of a similarity betweenCassel's and Milton Friedman's interpretations of the 1929 crisis). The contractionary bank policy of the Federal Reserve Banks was the main factor behind the deflationary process. It should be noted that in his discussion of excess investment in the 'usual' crises, Cassel (1932:625, 642) referred to the 'action of banks' as a secondary factor that only 'accentuates' the boom. Generally speaking, he was critical of Wicksell's view that a cumulative inflationary process would result from a bank rate set below the 'natural' rate, since the ensuing forced saving would increase the capital stock and bring the natural rate into equality with the market rate (see Cassel 1932:437 – 8; and Wicksell's reaction [1919:251]; see Lindahl 1939: 181 – 2, who accepted Cassel's objection under the condition that price expectations are not elastic; it is worth noting that Cassel's view that 'forced saving' constitutes an equilibrating factor sets him apart from the Austrian approach to economic crises).

Cassel often pointed out that the study of 'economic dynamics' should be essentially 'inductive', in contrast with the 'deductive' methods employed in 'general price theory' (see 1925:27 – 9; 1932:533 – 4).[19] He carried out an extensive empirical investigation of business cycles in the period 1870 – 1914 in Book IV of his *Social Economy,* where the historical interaction between changes in the production of new capital goods and oscillations of the interest rate was pointed out. It is worth noting that a certain degree of nominal wage rigidity was among his empirical results (1932:607), but he used that not as a factor contributing to the intensity of depressions, but as a necessary element in the explanation of how a persistent excess supply of labour could be consistent with positive wages (*ibid.:* 357). Cassel's (1925: 27) distinction between statics and dynamics prompted Myrdal (1927:11, quoted and translated by Hansson, 1982:33; this is Myrdal's Ph.D. thesis, written under Cassel's supervision) to construct a 'third, deductive, approximation' (to Cassel's stationary and steady states), that is, the equilibrium price formation of a society with non-uniform changes'. Cassel (1937:440) hailed 'recent endeavours to extend economic investigation beyond the case of equilibrium here referred to', which would lead to the building up of a 'coherent theory of dynamic economics by co-ordinating the theories of prices, money, and business cycles'. The study of the uniformly progressive economy remained, though, in his opinion, the necessary foundation of theoretical economics.

ACKNOWLEDGEMENTS

I am grateful to Daniele Besomi and Geoff Harcourt for helpful discussion. I also would like to thank Gilles Dostaler and other participants at the Annual European Conference on the History of Economics (Lisbon, 8 – 10 February

1996) for their comments. Financial support from the RoyalEconomic Society, St Edmund's College and the Faculty of Economics and Politics (Cambridge) towards the costs of attending the conference is gratefully acknowledged.

NOTES

1 It should be noted that the capital-output ratio is given in Harrod (1939, 1948) not for technological reasons, but because the market rate of interest is assumed to be constant. Moreover, as Kregel (1980) and Besomi (1995) have pointed out, the description of Harrod's as a steady state model is incorrect.

2 See Hicks 1985:9; Schumpeter 1954:966.

3 Cassel (1932:468 – 70) used the fact that the price level was the same in 1850 and 1910 to infer that the increase of the world gold stock (2.8 per cent per year) during that period corresponded to the increase in the demand for gold caused by economic growth. At 63 n2, he used data about the annual increase in the production of pig iron (4.2 per cent) and foodstuffs (1.2 per cent) to find—assuming that agriculture represented 1/3 and industry 2/3 of aggregate output—an annual rate of growth of 3.2 per cent.

4 Cassel (1935:127) referred also to increasing returns to explain the 'steady increase in efficiency', which is consistent with uniform growth (cf. Hahn and Matthews 1964:833).

5 It is worth noting that this criticism of classical economists—that is, the notion that the whole product is paid by consumers—formed the basis for Marx's development of his reproduction models, which might have been, besides Marshall, another source of influence on Cassel's concept of steady growth (see Marx 1992:509 – 12). The view that, even in equilibrium, savings are represented by stores of goods still survived in the business cycle literature of the early 1900s (Boianovsky 1955).

6 Cf. 1932:62—'The total income I *thus* stands in an invariable ratio to the total capital C' (emphasis added). Brems (1986:263) omits in his quotation of that sentence the crucial 'thus'. Cassel *(ibid.)* considered also the possibility that a growing capital-output ratio could be compensated for by a rising s, but dismissed it in the end (cf. Robinson 1952:45 – 6).

7 As Cassel (1932:39 – 40) put it, 'the existence of a certain amount of real capital at the initial point—and any given moment can be regarded as the initial point—is a necessary condition of the progressive state'. Instead of assuming $nK=sY$, Solow (1956:164) writes the general equation $\dot{K} = sF(\dot{K}, L_0e^{nt})$ for the equilibrium (but not necessarily steady state) path of the capital stock. It is worth noting that Cassel (1935:125 – 6) reacted against the extension of the notion of marginal productivity to an aggregate economy treated like a one-good economy, on the grounds that the aggregate production function could only be conceived of if either the different goods 'are combined with one another in fixed quantitative relations', or their prices, and, consequently, the factor prices, are known, which means that marginal productivity cannot be used to find those prices in the first place.

8 Cf. Hennings (1990) on the related—but distinct—concepts of 'waiting' and 'abstinence' put forward by Cassel, Marshall and Nassau Senior. It should be noted that Cassel (1932:248) rejected Marshall's notion of interest as the payment for a

'real cost' and extended, in contrast with Senior, the concept of waiting to a stationary economy in connection with the maintenance of *fixed*—instead of circulating—capital. Curiously enough, Veblen (1904:212, n31) referred to Cassel's (1904) use of the equivalent Swedish term 'väntande', but wrongly translated it as 'pecuniary expectancy' and, consequently, misunderstood Cassel's explanation of the crisis.

9 Cassel never really incorporated capital and waiting into his equations of uniform growth in general equilibrium (see Wicksell 1919:226). It is possible to find in the literature two distinct ways to close Cassel's model: either by incorporating fixed capital as a production factor like Walras (see Kuenne 1963:463ff) or by assuming that inputs precede outputs in time like Von Neumann (see Brems 1989:167). In both treatments there is no place for waiting as a production factor, in contrast with Cassel's own not fully developed suggestions (see 1932: 210).

10 There are two 'traditions' in economic growth theory as far as the treatment of the saving function is concerned: the Cassel-Harrod/Domar-Solow/Swan chain, which assumes a constant s, and the Wicksell-Ramsey-Koopmans (W-R-K) chain, which approaches saving from the point of view of intertemporal utility maximisation. The two traditions produce the same results in steady state (as s is then necessarily constant), but outside the steady state the W-R-K result shows that capital accumulation proceeds at a (variable) pace such that the marginal utility of consumption per capita falls at a proportional rate equal to the difference between the rate of interest and the subjective rate of time preference (see Boianovsky 1998a, for Wicksell's verbal formulation for the case of constant population and his criticism of Cassel's neglect of the influence of the rate of interest).

11 It should be noted, though, as Yeager (1976) has shown in his plea for the substitution of waiting for capital as the production factor for which interest is paid, that the demand function for waiting is not immune to the 'Cambridge critique' about reswitching and capital-reversal (see Harcourt 1972). Yeager sustains, however, that, since waiting has both time and value dimensions, the dependence of the amount of waiting required in a given production process on its own price should not be considered a 'paradox'. Needless to say, these questions were alien to Cassel, who (despite realising some of the problems involved in the measurement of variables in the theory of distribution, as we saw in note 7 above) ignored also Wicksell's discussion of what is now known as the 'Wicksell effect' in capital theory. Cf. Wicksell 1919:244: 'We must always introduce a reservation for the 'the marginal productivity of capital' regarded as a sum of value. This I have explicitly proved in my writings, but Professor Cassel completely neglects it'.

12 Böhm-Bawerk, in a letter to Wicksell dated 16 June 1901 (in the Wicksell Archives, Lund), reacted strongly against what he called Cassel's attempt 'to bring interest on capital into a causal relation to progress'. Magnusson's (1991: 131) suggestion that Cassel's treatment of interest is 'totally static' is, therefore, hardly accurate.

13 In his 1904 article on economic fluctuations, however, Cassel (1904:58) used the case of unitary *sp* in a capitalist economy to argue, against the underconsumptionist view, that even if capitalists save the whole of their income the economy could still be in equilibrium with a rate of growth equal to the rate of interest.

14 Lundberg did not refer to Cassel in this connection, even though he had mentioned before at 183 Cassel's 'uniform progressive economy' and given at 185, n1 a

precise formulation of the rate of output growth in continuous time equilibrium. Lundberg's (ch. 9) well known 'model sequences' were designed to show that only under special circumstances will the economy develop according to a steady path. He shows in particular that in the case of a 'variable interest rate' (216ff) the continuance of an expansion requires that current gross saving isalways enough to finance the investment made in the previous period, which will happen only if excess investment increases annually according to the rate (1−s)/s (see 231 – 2).

15 Cassel (1932:641) also considered an alternative equilibrating mechanism, that is, the increase in the costs of producing new capital goods. The crucial difference, of course, is that no crisis must occur in this case, since the mechanism works through the effect on plans for producing new capital goods, not through the reduction in prices of new capital goods already produced. See Walras (1954: 293), and especially Robinson's (1952:47 – 8) explanation in terms of Marshallian long-period supply prices.

16 It is not clear whether Cassel had in mind changes in the level of population or in the rate of population growth, although the latter is the relevant one in his framework. Besides, strictly speaking, only neutral technical progress is consistent with steady growth, unless there are compensatory changes in the propensity to save. Cassel's picture of the economy is closer to a succession of virtual steady states determined by changes in the rate of growth of the effective labour force n.

17 Brems' (1986:265 – 6) suggestion that Cassel can be read as a Harrod-like instability caused by a self-accentuating divergence between the actual and warranted growth rates has no support at all from the textual evidence, though. One of the 'stylised facts', according to Cassel (1935:99 – 100) is the 'comparatively small' deviations from the uniform growth, 'a very remarkable fact for which we must find a definite explanation' (cf. Solow 1988:11).

18 See Boianovsky 1998b for Cassel's interpretation of the 1920 – 2 deflation. See also Carlson's (1993:182ff) discussion of Cassel's (1937) rejection of Keynes's notion of permanent unemployment and of his reaction against expansionary fiscal policy in the 1930s.

19 Mitchell (1969:432), of all people, apparently did not appreciate Cassel's distinction, since he suggested that the Swedish economist used statistical material to provide 'an inductive verification of his deductive theory', which was not the case.

REFERENCES

Besomi, D. (1995) 'From the trade cycle to the 'Essay in dynamic theory': the Harrod-Keynes correspondence, 1937 – 8', *History of Political Economy,* 27 (2): 309 – 43.

Blaug, M. (1985) *Economic Theory in Retrospect,* 4th edn, Cambridge: Cambridge University Press.

Boianovsky, M. (1995) 'Wicksell's business cycle', *European Journal of the History of Economic Thought,* 2 (2): 375 – 411.

——(1998a) 'Wicksell, Ramsey and the theory of interest', *European Journal of the History of Economic Thought,* 5 (1), 140 – 68.

——(1998b) 'Wicksell on deflation in the early 1920s', *History of Political Economy,* 30 (2), 19 – 75.

Brems, H. (1986) *Pioneering Economic Theory, 1630 – 1980,* Baltimore MD: Johns Hopkins University Press.

——(1989) 'Gustav Cassel revisited', *History of Political Economy,* 21 (2):165 – 178.

Carlson, B. (1993) 'The long retreat: Gustav Cassel and Eli Heckscher on the "New Economics" of the 1930s', in L.Jonung (ed.) *Swedish Economic Thought: Explorations and Advances,* London: Routledge, 156 – 94.

Cassel, G. (1900) *Das Recht auf den vollen Arbeiterstrag,* Göttingen: Vandenhoeck & Ruprecht.

——(1903) *The Nature and Necessity of Interest,* London: Macmillan.

——(1904) 'Om kriser och daliga tider' [On crises and bad times], *Ekonomisk Tidskrift,* 4 (1, 2): 21 – 35, 51 – 81.

——(1925) *Fundamental Thoughts in Economics,* London: Unwin.

——(1928) 'The rate of interest, the bank rate, and the stabilisation of prices', *Quarterly Journal of Economics,* 42 (Aug): 511 – 29.

——(1932) [1918] *The Theory of Social Economy,* revised edn, trans. S.Barron, New York: Harcourt, Brace.

——(1933) 'The problem of business cycles', *Skandinaviska Kreditaktiebolaget Quarterly Report,* January: 1 – 4.

——(1934) *Teoretisk Socialekonomi,* Stockholm: Kooperativa Forbundets Bokfrlag.

——(1935) *On Quantitative Thinking in Economics,* Oxford: Clarendon Press.

——(1937) 'Keynes' "General Theory"', *International Labour Review,* 24 (4): 439 – 44.

Dorfman, R. (1959) 'Waiting and the period of production', *Quarterly Journal of Economics,* 73 (Aug): 351 – 72.

Ellis, H. (1934) *German Monetary Theory (1905 – 33),* Cambridge MA: Harvard University Press.

Fisher, I. (1907) *The Rate of Interest,* New York: Macmillan.

Haberler, G. (1937) [1946] *Prosperity and Depression,* New York: United Nations.

Hahn, F. and Matthews R. (1964) 'The theory of economic growth: a survey', *Economic Journal,* 74 (Dec): 779 – 902.

Hansen, A. (1951) *Business Cycle and National Income,* New York: Norton.

Hansson, B. (1982) *The Stockholm School and the Development of Dynamic Method,* London: Croom Helm.

Harcourt, G.C. (1972) *Some Cambridge Controversies in the Theory of Capital,* Cambridge: Cambridge University Press.

Harrod, R. (1939) 'An essay in dynamic theory', in A.Sen (ed.) (1970) *Growth Economics,* Harmondsworth: Penguin, 43 – 64.

Harrod, R. (1948) *Towards a Dynamic Economics,* London: Macmillan.

Hayek, F.Von (1933) [1966] *Monetary Theory and the Trade Cycle,* New York: Kelley.

Hennings, K. (1987) [1990] 'Waiting', in J.Eatwell *et al* (eds): *The New Palgrave: Capital Theory,* London: Macmillan: 243 – 6.

Hicks, J. (1985) *Methods of Dynamic Economics,* Oxford: Clarendon Press.

Keynes, J.M. (1937) [1973] 'The economic consequences of a declining population', in D.Moggridge (ed.) *The Collected Writings of John Maynard Keynes,* vol. 14: 125 – 33, Cambridge: Cambridge University Press.

Kregel, J. (1980) 'Economic dynamics and the theory of steady growth: an historical essay on Harrod's "knife-edge"', *History of Political Economy,* 12 (1): 97 – 123.

Kuenne, R. (1963) *The Theory of General Economic Equilibrium,* Princeton NJ: Princeton University Press.

Kurz, H. and Salvadori, N. (1993) 'Von Neumann's growth model and the "classical" tradition', *The European Journal of the History of Economic Thought,* 1 (1): 129 – 60.

Lindahl, E. (1939) [1970] *Studies in the Theory of Money and Capital,* New York: Kelley.

Lundberg, E. (1937) [1964] *Studies in the Theory of Economic Expansion,* New York: Kelley.

Lundberg, E. (1967) 'The influence of Gustav Cassel on economic doctrine and policy', *Skandinaviska Banken Quarterly Review.* 1 – 7.

Magnusson, L. (1991) 'Gustav Cassel, populariser and enigmatic Walrasian', In B. Sandelin (ed.) *The History of Swedish Economic Thought,* London: Routledge: 122 – 40.

Marshall, A. (1890) [1990] *Principles of Economics,* London: Macmillan.

Marx, K. (1885) [1992] *Capital,* vol. 2, London: Penguin.

Matthews, R. (1959) *The Trade Cycle,* Cambridge: Cambridge University Press.

Mitchell, W.C. (1969) *Types of Economic Theory,* vol. 2, New York: Kelley

Myrdal, G. (1927) *Prisbildningsproblemet och Föranderligheten,* Stockholm: Almqvist & Wiksell.

Niehans, J. (1990) *A History of Economic Theory,* Baltimore MD: Johns Hopkins University Press.

——(1995) 'Multiple discoveries in economic theory', *European Journal of the History of Economic Thought,* 2 (1):1 – 28.

Pasinetti, L. (1974) *Growth and Income Distribution,* Cambridge: Cambridge University Press.

Patinkin, D. (1965) *Money, Interest and Prices,* New York: Harper & Row.

Phelps Brown, H. (1981) 'Sir Roy Harrod: a note', *Economic Journal,* 91 (March): 231.

Robertson, D. (1926) [1989] *Banking Policy and the Price Level,* New York: Kelley.

Robinson, J. (1952) 'The model of an expanding economy', *The Economic Journal,* 62 (March): 42 – 53.

Rose, H. (1969) 'Real and monetary factors in the business cycle', *Journal of Money, Credit and Banking,* 1 (1):138 – 52.

Samuelson, P. (1993) 'Gustav Cassel's scientific innovations: Claims and realities', *History of Political Economy,* 25 (3):515 – 27.

Schumpeter, J. (1954) *History of Economic Analysis,* New York: Oxford University Press.

Sen, A. (ed.) (1970) *Growth Economics,* Harmondsworth: Penguin.

Solow, R. (1956) 'A contribution to the theory of economic growth', in A.Sen (ed.) (1970) *Growth Economics,* Harmondsworth: Penguin, 161 – 200.

——(1988) *Growth Theory: An Exposition,* New York: Oxford University Press.

Spengler, J. (1969) 'Cassel on population', *History of Political Economy,* 1 (1): 151 – 172.

Swan, T. (1956) 'Economic growth and capital accumulation', *Economic Record,* 32 (Nov): 334 – 43.

Tinbergen, J. and Polack, J. (1950) *The Dynamics of Business Cycles,* London: Routledge and Kegan Paul.

Veblen, T. (1904) *The Theory of Business Enterprise,* New York: Mentor.

Walras, L. (1874; 1926) [1954] *Elements of Pure Economics,* Homewood: Irwin.

Wicksell, K. (1919) [1934] 'Cassel's system of economies', In K.Wicksell, *Lectures on Political Economy,* vol. 1, London: Routledge.

Yeager, L. (1976) 'Toward understanding some paradoxes in capital theory', *Economic Inquiry,* 14 (3):313 – 46.

4
ALFRED MARSHALL AND EVANGELICALISM

R. McWilliams Tullberg

INTRODUCTION

In two articles written in the 1930s, Talcott Parsons described how the themes of wants and their adjustment to activities, so central to Marshall's theory of economic progress, were reflections of the Protestant virtues of thrift, sobriety, temperance, self-awareness and control, etc., which would lead not only to the growth of output but to the gentrification of those involved. Parsons put this 'narrow-mindedness, hardly compatible with the ideal of scientific objectivity' (1932:220) down to evangelical bias, a vestige of Marshall's fundamentalist childhood (Chasse 1984:382). Fifty years later, J.D.Chasse advanced the view that Marshall's analysis of wants and activities is better understood in terms of the contemporary intellectual ideas of utilitarianism, Hegelianism and German historicism. Whilst it is true that Marshall incorporated into his system ideas which can be attributed to a number of philosophical schools, my belief is that his habits of thought and behaviour as well as his moral outlook, were essentially evangelical, and that when he claimed that the two great influences on a man's character were his everyday work and his religious ideals, he was in no way excluding himself (Marshall 1961: vol. 1,1).

A symbiotic and dynamic relationship existed between the English evangelical movement and the whole spectrum of nineteenth-century economic, social and intellectual ideas. It both changed and was changed by them, and as the century progressed, evangelicals learned to regard former vices as virtues and virtues as vices. It is therefore necessary to examine evangelicalism both in theory and practice so as to assess its impact on Marshall's vision of the possibilities of social and economic progress.

EVANGELICALISM IN THEORY

If religion is considered as a philosophy or way of life rather than as a system of faith and worship, then the religion which dominated the lives ofVictorians was that of evangelicalism, even where allegiance was professed to the doctrines of

anglo-catholicism, unitarianism, agnosticism or secularism. The English evangelical movement had arisen from an attempt to counter the 'nominal' Christianity of the eighteenth century, the hallmark of which had been 'a sort of formal, general assent to Christianity in the gross, and a degree of morality in practice' (William Wilberforce, quoted in Houghton 1974:228).[1] Evangelicalism was a restatement of the doctrine as to the essence of the Gospel message. Faith in the atoning death of Christ is the only means of salvation. The sacraments and good works have no saving efficacy *per se*. All human actions, even good works, are tainted by Original Sin, so it is impossible by this means to gain merit in the sight of God. Defined in this way, the essential evangelical dogma seems to bear little resemblance to the modes of thought and behaviour which enveloped nineteenth-century Britain. Even a casual acquaintance with the literature or economic, social and political history of the period reveals that much more was demanded of a respectable Victorian than a simple act of faith. Faith, it was preached, necessarily produced more holiness. Moral behaviour and good works were not only a prerequisite for salvation but proof of its occurrence, a mark of gratitude on the part of the saved and a method of nurturing the act of faith which had led to salvation. Thus the demands of piety, duty, judgement, sobriety, seriousness and Sunday-observance were inescapable and 'good works' (sometimes eccentrically defined) were required of all evangelicals. Every human action was subject to scrutiny. Though closely related in England with methodism, the term 'evangelical' was used by many non-conformist groups and 'chapel-goers', as well as to the Low Church Party of the Church of England.

The special dogmatic features of evangelicalism have been identified as conversion, activism, biblicalism and 'crucicentrism'—that is, emphasis on Christ's sacrifice on the cross.[2] Human beings are estranged from God by their sinfulness; salvation could only be received by faith and divine grace. This involved *conversion,* either gradual or as the result of a crisis. Conversion demanded *activism* on the part of believers. This meant missionary work at home and abroad, in sharp contrast to what was regarded as the 'torpor' of eighteenth-century religious practice. Certain sects required their members to 'witness' whenever they found themselves among the unconverted.[3] All evangelicals held the Bible in particular reverence. For some, every word of the Bible was literally true; others regarded it as the major devotional book and source of inspiration. Finally, emphasis was placed on the doctrine of the atonement and salvation through Christ's crucifixion, rather than on redemption through the incarnation.[4] This raised the question—for whom did Christ die? A major divide existed between the Calvinists who believed salvation was only for the elect, and those of the 'Arminian' persuasion who believed it was general, that is, open to all. The latter view became the dominant one in England and Wales.

EVANGELICALISM IN PRACTICE

As Edward Gibbon wryly remarked: 'Religion has never existed in the pure form in which it descended from Heaven' (quoted by Bebbington 1989:ix), and we cannot expect the social expression of evangelical belief to be more than loosely associated with evangelical theology in its purest form. Commentators have sought to explain the discrepancy between evangelicalism in theory and evangelicalism in fact. Boyd Hilton (1985:215 – 16) feels there was a significant correlation between the moral/social and theological thinking of the established church. This he attributes to changing interpretations of the meaning of divine providence as it made itself known on earth. Some groups, of which the Clapham Sect was the most famous, held that God operated through the immutable laws of nature (including economic laws as then understood) and these should not be interfered with. If God had created poverty, then it would be an affront to His wisdom to try to alleviate it. Other evangelicals believed in the theory of 'special providence' whereby God intervened on earth with punishments for spiritual and moral transgressions. Providence intervened 'in each twist and turn of a man's life' (Annan 1984:8), guiding the soul along the spiritual and moral road to God, and as such was unrelated to 'natural' laws. Such laws could therefore be flouted in the interest of good government which would foster self-improvement and stamp out occasions for sin. E.R.Norman (1976:10) argues strongly that 'most of the social attitudes of the Church have derived from the surrounding intellectual and political culture and not, as churchmen themselves always seem to assume, from theological learning', and that this was necessarily the case since the clergy were educated at Oxford and Cambridge alongside other members of the ruling and opinion-building classes, shared their values and lifestyles and were influenced by the same ideas.[5] Norman's and Hilton's arguments are not incompatible, since the former simply begins where the latter leaves off—what did evangelicals believe was 'good government'? The answer to this question they found in the intellectual and political climate imbibed by men of their own class and education.

Bebbington (1989), while siding with Norman, does much to cut through this dispute by defining nineteenth-century English evangelicalism not in terms of religious dogma but as a matter of emphasis.[6] Evangelicalism was all strands of Protestantism which were neither too high in churchmanship nor too broad in theology, and it spanned the gulf between the established church and non-conformity (Bebbington 1989:ix). It was not distinguished by any special doctrines exclusive to evangelicals but by the emphasis given to certain doctrines over others, and the view that these doctrines should mould not only the conscience but also the conduct of believers. Naturally, over time, shifts of emphasis could and did take place, and doctrines were reinterpreted to accommodate current intellectual fashions. However, certainhabits of thought and behaviour remained unchanged throughout the century and left a special imprint on the majority of Victorians.

The basic aim of evangelicalism, that of arousing religious and moral awareness to a degree where God was involved in a person's behaviour every second of every day, infected the attitude of the entire nation, irrespective of belief or disbelief, and evangelical respectability became the well-nigh essential costume of the upwardly mobile.[7] Reaction to evangelicalism must never be indifference; the greatest social sin, greater possibly than the holding of heretical opinions, became non-engagement or indifference. The pervasiveness of evangelical ideas is stressed by many commentators; Houghton (1974: 238) for example, notes that 'the creed of earnestness pushed its way beyond the church walls into the community at large. Its ideals penetrated into the homes—and consciences—of half-believers and outright agnostics'. Annan described evangelical morality as: 'the single most widespread influence in Victorian England. It powerfully influenced the Church of England, was the faith of the Methodists, and revived the older non-Conformist sects, it spread through every class and taught a clear set of values' (Annan 1951:110).

Evangelicals had a very special awareness of God. In particular, they did not allow salvation through Jesus Christ to cloud their appreciation of God as supreme arbiter: 'though the atoning power of Christ's sacrifice was the central doctrine of Christianity, God's judgement—indeed His duty to judge—remained the final reality' (Annan 1951:112). This led to a special awareness amongst evangelicals of the need to judge between right and wrong, both in their own behaviour and that of others. Even dissociated from its religious foundations, judgement could not be suspended. The fear of moral relativism which would accompany intellectual radicalism inhibited many would-be non-believers from acknowledging their doubts and led others, like Marshall, to claim that while they rejected theology, they still believed in religion, by which they meant not only the consolation it might provide but also its moral discipline.

The evangelical movement was intensely puritan. Tremendous importance was attached to self-denial, the rejection of pleasures, the curbing and training of the will. Major vice could only be resisted in adulthood if one had been taught in childhood to overcome minor bad habits. By most accounts, evangelical morality weighed heavily on children. Judgement could not be left to their immature consciences, nor could the search for a personal knowledge of God be left to their untrained minds. This also applied to those with 'weaker' characters and intellects—women, the lower classes and other races. Their parents, guardians, husbands or masters were to be their teachers, and at times sadism was combined with religious fervour with frightening results. Criticism also extended to the 'idle rich', not least because they had the time, the means and, it was assumed, the mental capacity to understand and practise the precepts of evangelicalism and to set a good example to the lower ranks of society.

A particular problem of evangelicalism in its early phase was its anti-intellectualism. The acquisition of knowledge would not bring a soul to heaven. Indeed, it might even make a person proud and lead them astray; at the very least

it would be a distraction. This in turn led to the condemnation of behaviour or pursuits which might *be* the source of corruption of souls. The fear that something might be an occasion of sin led to blanket bans on literature and entertainment without any real analysis of individual cases. 'Fiction made light of sin, wasted time and had worldly associations' (Bebbington 1989:131). The same was said of art, music and even sports.[8] Even those more enlightened souls who could see no wrong in these things *per se,* recognised them as wasting time which could be used in devotional study, self-improvement and proselytising.

One of the more unpleasant results of evangelical practice was the guilt which became associated with enjoyment and relaxation. Marshall himself offered a typical example of the inhibitions created. Having initially permitted chess, his father made him promise never to play it again. Marshall kept his promise all his life, but never failed to be excited if he saw a chess problem in a newspaper. However, Marshall felt his father had been right to exact the promise; otherwise he would have been tempted to spend all his time on chess. He reportedly said, 'We are not at liberty to play chess games, or exercise ourselves upon subtleties that lead nowhere' (Keynes 1985:163, quoting Mary Paley Marshall). In other words, there was nothing morally suspect about chess, but it wasted time which could be used for 'better' purposes.

The dilemma was only partially resolved later in the century by giving these worldly pursuits a moral twist; novels could contain a suitable Christian message, music and art a religious motif, and football became an ingredient of 'muscular Christianity'. What early nineteenth-century evangelicals saw as worldly came to be recognised as having a potential in the churches' mission. But it was still necessary to justify pleasurable pursuits; much of the guilt remained.

In a similar fashion, the early aversion of evangelicals for politics as wicked, corrupt and worldly gave way to an understanding of the ways in which politics could be used to combat sin and further the evangelical cause. Bebbington identifies three broad classes of evil which stirred evangelicals to political action. These were:

1 obstacles to spreading the gospel;
2 substitutes for the gospel, i.e. 'false' religions; and
3 sin, by which was usually meant sex (Bebbington 1989:133).

He argues that it is a mistake to describe evangelical crusades as motivated primarily by humanitarianism, since they were regarded as attacks on sin and evil which evangelicals were morally bound to fight *(ibid)*.

Thus Bebbington, though a sympathetic historian of two centuries of evangelicalism, sees little evidence of any sense of justice or civil liberties in evangelical campaigns. The most famous of these, the anti-slavery movement, began because slave owners opposed missionaries in the Caribbean; factory acts were not prompted by the inhuman conditions under which women and children

were working in mines, but rather by their nakedness during working hours. Concern with overcrowding in the slums was not a reaction to the stress and distress of slum life but to a realisation that living so tightly packed together led to 'immorality' in a variety of forms, including incest. Drunkenness and gambling, the next two deadly sins, not only 'deadened the senses' and shut out the word of God but could also lead to immorality. 'The dynamism of the Evangelical approach to politics was hostility to sin' (Bebbington 1989:135).

Within one sphere, however, the evangelical doctrines offered encouragement, and it would be wrong to paint a picture of evangelical practice as totally other-worldly. Many who were taught the precepts of restraint, self-improvement, application, duty, discipline and perseverance kept their religious observance for Sundays. During the rest of the week, they applied the precepts in the cause of the second arm of their Victorian faith—the possibilities of material progress.[9] This is not to suggest that evangelical doctrines specifically promoted the habits of accumulation on which material progress was based. By comparison with other religions which tended to emphasise a 'bank book' approach to salvation, whereby good works and self-mortification in this world accumulate as a positive balance for use in the next, the Protestant ethic into which evangelicalism sought to breathe new life stressed the meaninglessness of such activities in the quest for salvation. Yet the sober and reflective behaviour encouraged by evangelicals in conjunction with the 'commercial spirit [which] has always existed in society' (Houghton 1974:183) produced a formula for commercial success until the latter became confused with the goal of eternal salvation itself.

This distortion of evangelical ethics became even more pronounced as evangelicalism won territory within the established church and became the badge of middle-class respectability.[10] Material wealth became another proof of sanctifying grace, and since the new morality placed narrow limits on the ways in which it could be used, it was put to work to produce more wealth, or, as the century progressed, to finance a game with an infinite subtlety of rules, the process of upward social mobility.[11]

The moral approval of wealth, wisely accumulated and spent, also produced a new attitude to poverty. Previously regarded as an inevitable element in society to be pitied and, out of Christian charity, relieved, it was now regarded as a sign of, at best, thoughtless living and a failure to attend to duty, and at worst, of moral depravity and brutishness which the victims had brought upon themselves. Most of all, it was despised because it was so greatly feared by those still climbing the infinitely graduated social ladder.For many, the fear of losing social status became more real that the fires of hell. Subconsciously, the real causes of poverty—'irregularities' of trade, unfortunate investments, death or illness in the family—were well understood, and a line was therefore drawn between 'deserving' and 'undeserving' poor, so that the class distinctions could be maintained should economic disaster strike. Every Victorian novel had

characters who 'kept up appearances' in the midst of 'genteel poverty', and others who through weakness of character sank into some temporal (and most probably eternal) Hades. Others continued to justify poverty as ordained by divine providence as described in the verse of the much-loved hymn, *All Things Bright and Beautiful:*

> The rich man in his castle,
> The poor man at his gate,
> God made them, high or lowly,
> And order'd their estate.
>
> (Cecil Frances Alexander, 1848)

This worldly version of evangelicalism did not go unchallenged by its opponents both inside and outside the established church. Contemporary prophets of a kinder, more caring society attacked not so much the church but the economists who 'justified' such a cruel interpretation of the individual's struggle on earth. Economists were accused of teaching that the mathematics of God or nature were such that the wages of the working classes would decline to subsistence level and the profits of employers would be soaked up by the land-owning classes. Welfare was forgotten in the struggle for wealth, a slice of the cake before it disappeared into someone else's pocket, along with such concepts as the fair wage, the just price and a reasonable level of profit. In the mid-Victorian reappraisal of doctrines, sacred and secular, it was therefore found necessary to reinterpret the life of the spirit both eternal and temporal.

MID-CENTURY REAPPRAISAL

A church attendance census of 1850 revealed that just over 50 per cent of the population over ten years of age attended places of worship. The Victorians were shocked by what they regarded as a 'poor' level of church attendance. In particular, the returns for inner-city working-class districts were much lower than those from country areas. This led to calls for renewed religious fervour and for special measures to bring the evangelical message to the urban working classes.[12] The result was a 'second revival' in the 1850s and 1860s, during which period the influence of evangelical values was at its peak.

The evangelical revival coincided with, and was perhaps made possible by, a shift of emphasis in evangelical thought away from a world of sinners, alienated from God by guilt and depravity, who could only be saved from eternal punishment through Christ's sacrifice on the cross, to a growing interest in Christ's humanity and compassion and his teachings in the New Testament. The transformation was due to biblical criticism which challenged the accuracy of many of the Bible stories, to scientific advances which challenged the earlier static and pessimistic view of the world as 'a vale of tears', to the relative

political stability of the period, and to the cult of 'refinement' induced by evangelicalism itself, so that gruesome descriptions of hellfire and the sufferings of Calvary involved too much physical detail for delicate Victorian susceptibilities.

At a fundamental level, the concept of providence altered from that of grim, unavoidable natural laws to that of retribution for specific faults which we must correct (see above).[13] Among some sections of the church, 'caring' for one's brother became less judgemental and paternalistic and aimed instead to understand and encourage. The evangelical virtue of earnestness had now to be coupled with enthusiasm for solving social problems. In time, it was realised that many problems had neither a spiritual cause nor solution. At its most intelligent, evangelicalism was a generous, non-authoritarian faith which gave great freedom to the individual in the pursuit of truth; indeed, it bore the seeds of its own destruction, since it led many inquirers to go 'too far' and ended with their secession from the faith.[14]

The rejection of religious beliefs was not an easy process. The mental and physical torments of doubt are well documented in both the fiction and biography of the period. Doubting the existence of a father-god and a life after death was castigated as a sign of intellectual arrogance and moral turpitude. One escape was the flight to the disciplines and certainties of Rome; another alternative was the wilderness of agnosticism.[15] Both solutions often involved a distressing break with family and friends, but the latter also meant the loss of a source of comfort and, most terrifying of all, the loss of moral guidance. The third escape route, and that taken by Alfred Marshall, was to drop the metaphysics quietly and privately, but to keep the rest of the evangelical baggage.

For many Victorians, the sole reason for their clinging to formal religion was their fear that its demise would necessarily involve the moral collapse of society.[16] Victorians associated the term 'moral collapse' with the civil tension and disorder of Chartism, the course of events in revolutionary Europe, and later in the century, the new trade unionism and socialism. The particular need to guard against social unrest coloured solutions to the problem of life with or without God. As J.S.Mill wrote in his *Autobiography:*

> The world would be astonished if it knew how great a proportion of its brightest ornaments—of those most distinguished even in popular estimation for wisdom and virtue—are complete sceptics in religion; many of them refraining from avowal, less from personal considerations, than from a conscientious, though in my opinion, a most mistaken apprehension lest by speaking out what would tend to weaken existing beliefs, and by consequence (as they suppose) existing restraints, they should do harm instead of good.
>
> (Mill 1873:45)

Amongst believers this was a further reason to increase their activities concerning both the temporal bodies and immortal souls of the poorest classes: anxiety to administer a prophylactic against social disruption was concealed behind the slogans of Christianity. Non-believers who could not base concern for their fellow men on the doctrine of the redemption, felt their responsibility the more greatly since the anticipated social unrest was blamed on their own rejection of God.[17]

To assuage their fears and to satisfy the needs of both believers and agnostics, the new faiths had to be shown to be as dogmatic as the old, built on the writings of new evangelists.[18] As Matthew Arnold's niece, Mrs Humphrey Ward, put it in her three-volume bestseller on the struggles of the modern religious conscience: 'The problem of the world at this moment is—*how to find a religion?* Some great conception which shall be once more capable, as the old were capable, of welding societies and keeping men's brutish elements in check' (Ward 1888:354 – 5). The search was conducted with the utmost fervour by men who demonstrated 'unparalleled vitality' when released from the yoke of narrow religious beliefs which repressed all mental and physical pleasures as sin (Smith 1967:6).

However, active agnosticism was something for the intelligentsia and the loss of faith must not be exaggerated. Smith, in his book *The London Heretics,* has pointed out that:

> From a purely external viewpoint, traditional religion did not seem disturbed by these intellectual inundations [evolutionary theory and biblical criticism]. Chapels and churches remained, for the most part, crowded. Books of conventional sermons were always among the best sellers. Missionary zeal was high. The American revivalists, Dwight L.Moody and Ira D.Sankey stormed London in 1875 after a wildly successful year in other parts of the British Isles.
>
> (Smith 1967:3)[19]

It was still possible for religious controversy to become a matter of intense public debate, and the opinions of clerics were referred and deferred to on many issues.

RELIGIOUS ISSUES AT CAMBRIDGE

The period of mid-century religious reappraisal coincided with Alfred Marshall's childhood and young adulthood. At home and at school he had experienced 'old-fashioned' evangelicalism, but as a young academic at Cambridge he was thrown into a maelstrom of conflicting views, not only on the practice of evangelicalism but also on its metaphysical foundations.

Green (1964: ch. 10) notes that although neither Oxford nor Cambridge can be considered the 'home' of evangelicalism, Cambridge had become closely associated with evangelical teaching in the early nineteenth century and, under the guidance of Charles Simeon (1759 – 1836), was something of a forcing house for evangelicals, lay and clerical.[20] In the second half of the century, a significant change in attitude towards religious matters occurred at Cambridge, although it is necessary to retain a sense of proportion as to the extent of agnosticism at the university.

The outward signs of the change were the repeal of the Test Acts in 1871, and the decline in candidates for the priesthood.[21] The year 1870 is frequently offered as the date after which it can be said that opinion was 'decisively secularised' (Annan 1951:150). Smith also sees agnosticism as becoming 'active' after 1870.[22] The intellectual catalysts had been German biblical criticism, the work of the natural scientists, in particular Darwin's *The Origin of Species* (1859) and, perhaps most important of all, modern interpretations of Christian theology.

Before Marshall became an undergraduate, the issue of the historical truth or falsity of Christianity was being hotly debated in Germany. The Bible was examined as a work of history and found wanting. Subject to textual criticism, it was found to be the work of many hands, at many dates. Some argued that this new scholarship did not alter Christian dogmas—but it was a hard blow for those who had believed in the literal or near-literal truth of the scriptures. In 1859, H.L.Mansel, an Oxford don, published his controversial Bampton Lectures, *The Limits of Religious Thought,* influenced by a combination of Kantian philosophy and Scottish intuitionalism favoured by Sir Walter Hamilton. In these he argued that God is infinite and unknowable; nevertheless our intuition tells us that he exists and that we are dependent on him. Our sense of moral obligation stems from the same source. The purpose of philosophy is not to arrive at a knowledge of the absolute nature of God, but 'to teach us to know ourselves and the limits of our faculties' (Mansel 1858, quoted in Oppenheimer 1988:121).

The lectures were attacked by the rationalists who regarded them as metaphysical mumbo-jumbo. John Stuart Mill, the leading exponent of the scientific method of investigation, replied to Mansel in his *Examination of Sir William Hamilton's Philosophy* of 1865, the year of Marshall's graduation. Mansel responded and the ensuing controversy caught Marshall's attention. If Keynes is to be believed, Marshall's reaction to Mansel'sinterpretation of religion —his 'defence of orthodoxy'—was that it 'showed me how much there was to be defended' (Keynes 1985:169). In his draft preface to *Money, Credit and Commerce* (1917) Marshall fleshes out this rather enigmatic statement, writing that the *Bampton Lectures* 'caused me to think that man's own possibilities were the most important subject for his study'.[23]

Like many others in Cambridge, Marshall plunged into the study of ontological and epistemological issues—what is the nature of man and how is he distinguished from beasts; how can he attain knowledge of himself and the

surrounding world? Marshall claimed that this was Mansel's doing.[24] But we must be careful in assigning this newfound interest in philosophical issues solely to doubts about religious faith. During three and a half years of study for the mathematics tripos, Marshall had been drilled in another faith, the axiomatic truth of Euclidean geometry:

> The study of geometry demonstrated the existence of absolute truth, capable of being reached by pure thought. In this one field, scholars believed that they had penetrated to absolute truth by the action of pure reason laying bare the structure of the universe…. The geometry of Euclid assumed that there were certain axiomatic truths, which did not have to be proved by empirical observations; they were simple, self-obviously, true.
>
> (Butler 1990:9)[25]

The Mansel/Mill controversy was provocative of new thinking in a number of fields, not just religion—what were the limits of our faculties and what lay beyond? Must we, as the rationalists insisted, stop at these 'limits' or could we speculate beyond them?

Another controversial and influential book, *Ecce Homo,* a study of the humanity of Christ, was published in 1865 by J.R.Seeley.[26] Although it resembled similar continental works, it was written for the leaders of British society, defining their moral responsibilities in a modern industrial society. Seeley was critical of the traditions and morality of the ancient world which he condemned as 'obsolete'—a viewpoint which appealed strongly to Marshall—and highlighted instead the flexibility and viability of the Christian value of the brotherhood of man (Rothblatt 1968:159 – 60). Thus he sought to bring to an end needless discussion on points of dogma and organisation, and to promote discussion on the role of the Christian morality in the light of contemporary scholarship in history, philosophy, psychology and science:

> Freed from pointless controversies, finally realising that the maxims of Horace had no effect applied to such new vices as railway profiteering and adulterated goods, the Church and its clergy would beable to address themselves to real and critical problems arising from the changing nature of English social and economic life.
>
> (Rothblatt 1968:161)

In January 1865, the drudgery of learning propositions and problem-solving techniques for the mathematical tripos finally over, Marshall was able to turn his mind to philosophical enquiries.[27] He became a member of a discussion group, the Grote Club, and on his own account began to study the Mansel/Mill debate in 1867 or thereabouts (Keynes 1985:166 – 7, 171). From ontological and epistemological issues he moved on to 'what seemed to be the more progressive

study of psychology' (Keynes 1985:171), or more specifically to ethics. During his psychological/ethical studies, he was faced with questions concerning the absolute and relative availability of material resources and their distribution, and was advised by 'older and wiser' men to study political economy. This, according to Keynes' *Memorial* (Keynes 1985: 179) was also in 1867. However, Keynes tells us that in 1868, 'when still in his metaphysical stage', the desire to read Kant in the original led him to Dresden to study German. In the same year (1868) he was appointed to a special lectureship in the moral sciences at St John's College 'and soon settled down to Economics, though for a time he gave short courses on other branches of Moral Science—on Logic and on Bentham' (Keynes 1985:172).

Chronologically, there is something wrong with the Keynes/Marshall version of Marshall's early intellectual development as a progression from mathematician considering a career in molecular physics, via metaphysics, psychological and ethical studies, to a social scientist specialising in economics.[28] In the full version of the recollections of Grote Club meetings which Marshall wrote for the Henry Sidgwick *Memoir* (A.S. and E.M.S. 1906) Marshall begins by explaining that in 1867 when he joined the Grote Club, he was a beginner in philosophy: 'I had only begun to read it seriously towards the end of 1865 and had been teaching mathematics all the while' (Add Ms.c.104[65] HSA). Yet in the same year as he joined the Grote Club (1867) Marshall began his study of political economy (Keynes 1985: 179 – 80); Whitaker supports this date, and it seems to tie in with Marshall's own recollection, given in a letter to James Ward in 1900: 'When Pearson asked me to lecture on Political Economy I consented; but I should have preferred philosophy, which was his subject. Shortly after [1868], the College made me a lecturer; and I added Logic and Ethics' (Pigou 1925:418).

In this same letter to Ward, he says that it was not until 1871 – 2 that he finally gave up psychology for economics (Pigou 1925:419). Further, it was after his interest in economics had developed to a point where he could lecture in the subject that he visited Germany (1868), 'still in his metaphysical stage' (Keynes 1985:172) to study German so as to read Kant in the original. Thus Marshall developed along a number of routes simultaneously,and in no sense dropped metaphysics and psychology in order to concentrate on economics but kept them in his baggage along the way.

In a similar fashion, I believe it can be argued that, unlike the received version of the story of Marshall's early abandonment of his faith, it remained integral to his economic vision. As an undergraduate, Marshall is said by Keynes to have looked forward to ordination and to work in the foreign missions; then, after 'a quick struggle, religious beliefs dropped away' (Keynes 1985:167). No evidence remains from the period in question; Marshall did not feel obliged to make public his loss of faith. Unlike Henry Sidgwick, Leslie Stephen and others who gave up their college fellowships on the grounds that they could no longer

subscribe to the thirty-nine articles of faith of the established church, Marshall kept his fellowship at St John's until he was obliged to relinquish it on marriage in 1877. Nor did he use the opportunity offered by his appointment in 1868 as college lecturer to give up his fellowship.[29] He campaigned for the abolition of the religious tests, but then so did many others who were themselves devout churchmen. As he later explained, he was in no way antagonistic to supernatural religion, but held that the documentary evidence in support of such religion was weak (Marshall to Benjamin Kidd, 6 June 1894, Marshall 1/72, MLA). Marshall's loss of faith happened gradually and never completely, and it remained a great source of inspiration to him. He was also much influenced by new thinking within radical evangelicalism. At an early stage, Marshall was drawn to the 'condition of the people' question, possibly by hearing Maurice, Kingsley and Fawcett speak, though more directly by association with the younger Grote Club members, in particular Henry Sidgwick and his circle.[30]

The history of upper- and middle-class concern with the lives of the lower classes in the second half of the nineteenth century is similar in broad outline to the concern for Third World poverty which grew among the richer nations in the second half of the twentieth century. After the Second World War, there was a reassertion of the western-style, culturally Christian value of caring for one's fellow men, irrespective of race and creed. The movement was initially led by intellectuals and fuelled by the growth of international news coverage. Whereas interest had previously centred on the salvation of foreign souls, in the postwar period both religious and secular groups began to respond in a more organised fashion to economic deprivation. They depended entirely on voluntary contributions and concentrated most often on emergency relief following natural and human-made disasters. In a second phase, the need for repeated doses of, emergency relief led, on the one hand, to a growing emphasis on non-emergency 'help-to-self-help' projects and to greater coordination and cooperation among the aid organisations; and on the other, to the involvement of governments of the richer world in the problems of Third-World poverty. In addition, it fuelled the study of 'development economies'—what are the mainsprings of growth?—Why had growth failed to occur in so many countries?—Why 'them' and not 'us'? The third phase has been an economic-political one, addressing questions of rights, justice, exploitation and colonialism, as well as international stability, economic expediency and economic self-interest. In the first two phases, a great deal of effort was put into trying to make 'them' like 'us'. Donors are now trying to learn that 'they' may not want to be like 'us', but that this is no justification for injustice and exploitation. These three phases run concurrently, and while the research and talking continues, individuals go on responding to the immediate needs of the starving child by dropping their coins into a collecting box, ignorant of or impervious to the wider issues, or impatient of their solution. In general, at least among private individuals, there has been real sincerity of purpose.

The nineteenth-century parallels are obvious, the only difference being the balance between the secular and the religious involvement. Marshall appears on the stage as a young Cambridge graduate at the equivalent of phase two. He was caught up in the vision of help-to-self-help and also in the study of economics to find the solution to the riddle of poverty. Any tendency to shudder at Marshall's sanctimonious 'priggishness' during the 1860s and 1870s can perhaps be resisted if we consider how our own attitudes to 'undeveloped countries' in the 1960s and 1970s will be judged a hundred years hence. Marshall's failure in the decades which followed to develop his understanding of the fundamental problems of poverty are less admirable, and form the subject of another paper.

Marshall first felt he had to deal with the issue of economic laws—were they fixed, as in the earlier version of the doctrine of providence, or were they scientific generalisations offering guidance on problems of moral behaviour (see above)? Although Hilton (1988:5) and other commentators describe the change of emphasis in evangelical thinking to that of 'special providence' as more or less complete by 1870, Marshall felt the message needed repeating. The case for the new *contra* the old version of providence is stated quite specifically in his 'Lectures to women' notes of 1873, and is referred to indirectly for the rest of the decade:

> An account [is] sometimes given of which it is difficult to speak temperately *viz.* 'it is a merciful dispensation that will, we trust, always prevent the cultivation of the lower classes. Some persons must be doomed to toil and debasing occupations: to give them culture would only be to make them miserable: to strain against this is to fly in the face of Providence: do not enquire into the cause; accept with gratitude the beneficial arrangement and pass on'.
>
> Yes: it is an arrangement of Providence and so is the cholera; *yet we may try to avoid it:* and…who dares to say thank God that I am comfortable and more or less cultivated and that others aredebarred from the higher forms of existence in order that I may thrive in the full plenitude of my might?
>
> (Raffaelli *et al* 1995:141, emphasis added)[31]

Marshall impressed his audience 'with the necessity of acting, with the special circumstances which render it impossible for us to look around us and yet be willing to drift' (*ibid.:* 125).

Two possible reasons exist for this emphasis on action. First, in this early stage of his career as an economist, Marshall may have felt the need to convince himself at least as much as his students of the propriety of interfering with 'economic laws'. The second possible reason is concerned with university politics. One leading exponent of the 'modern' view of providence was F.D.Maurice, who had lost his professorial chairs at King's College London in 1853 because of his association with radicals (leading to a charge of infidelity)

and his rejection of the doctrine of eternal damnation as 'superstition'. He became Knightsbridge Professor of Philosophy in Cambridge in 1866, that is, during the early days of Marshall's search for a niche in the moral sciences. Though a sick man, Maurice's appointment provided moral support for university radicals. On his death in 1872, the university made another controversial appointment to the philosophy chair, this time of the ultra-conservative Thomas Rawson Birks, 'who believed strongly in prophecy, inspiration, pre-millennialism and eternal punishment of the wicked' (Hilton 1988:365).[32] Alfred Slater West, Senior Moralist in 1869 and an internal examiner for the moral sciences tripos, in analysing the difficulties hindering the development of moral science subjects at Cambridge, cited as chief impediment the 'spirit of theological dogmatism'. In particular, he described Birks' inability (or unwillingness) to distinguish between 'Laws Proper of Jurisprudence and Laws metaphorically so-called (*i.e.,* scientific generalisations)' (West 1874:456). Cambridge men may therefore have had to continue fighting battles which others felt were already won.

Marshall made good use of the language of religion in his series of six 'popular lectures to women' in 1873;[33] the symbolism was well understood by his audience and could prompt the desired habits of thought and practice.[34] He told his audience that God had chosen 'ladies' to do the work of developing in the proletariat (Marshall's choice of terminology) the 'noble sources of joy' which he identified as cleanliness, independence, a sensitive conscience, a job properly done, responsibility for one's property and one's children (Raffaelli *et al.* 1995: 118 – 19). And his audience would have had no difficulty in recognising the essentially evangelical view of life—praising God through good personal habits, care of one's property and family, and hard work. Marshall also managed to weave into his lecture his dislike of exaggerated study of the language and culture of the classical Greeks and Romans. They were grossly pagan, immoral and reckless of human life.They were dilettantes and lacked the moral enthusiasm appropriate to Christians. Thus Greeks and Romans believed slavery was natural and the modern world had followed them, saying that the proletariat was necessary and natural. We should learn from the intellectual achievements of the Greeks and Romans but not from their ethics. Remember we are 'followers of a person who maintained with the most unflinching audacity of doctrine that a proletariat was not necessary'. Christ condemned the enslavement of men to machines, and so do political economists (Raffaelli *et al.* 1995:126).

Marshall was horrified by the waste of intellectual capital which the use of men instead of machines or the use of men to the point of exhaustion involved, and reasoned that they could never use even their limited leisure time for self-improvement if they were exhausted. If hours of work could be reduced, workers could be cultivated to the level of gentlemen. They would then stop contracting imprudent marriages and besotting their minds with drink, and would pay for their children's education (Raffaelli *et al.* 1995: 105 – 6). The benefits of educating children would, Marshall admits, mainly accrue to the nation as a

whole and to coming generations, and a case could therefore be made for state subsidisation of education. However, although the lower classes will rarely make 'the necessary sacrifice' to educate their children (*ibid.:* 103 – 4, 107), it would be a great mistake to diminish the responsibility of parents in this matter—'it does more harm than any good can compensate' (*ibid.:* 104):

> if we send a child to school, we do good; but if we induce his father to send him we do much more good; we have opened up to the latter new sources of joy, have given him new tendencies to work for an end. In this way we do a good the end of which we cannot see.
>
> (*ibid.:* 127)

Those with adequate incomes also had responsibilities and were not free to spend their money 'gratifying their appetites and taste for luxuries' (*ibid.:* 103). Expenditure on culture such as a box at the opera, an art education or travelling, etc., creates capital in the sense of stored up sources of enjoyment, and is therefore, subject to limitations, not wrongful. It would become wrong if, earning money in order to finance these luxuries, they wasted time, which is the most valuable capital of all. Fashionable dress and other means of display are vicious because they create the pains of envy in others (*ibid.:* 123 – 4).

These normative ideas appear repeatedly throughout Marshall's writing, though their presentation reflects his growing awareness and final assurance of his position as the country's leading economist. The 'Preliminary survey' in the *Principles* gives a good example of this. The noble sources of joy—cleanliness, responsibility, self-sufficiency, etc.—reappear as the virtues of the steadfast British workman, giving him the cutting edge over other races.Slavery was a feature of pagan times; Christianity has proclaimed the dignity of man and has called into question the very existence of a lower class 'doomed from their birth to hard work in order to provide for others the requisites of a refined and cultured life'. These people are subject to exhausting toil, malnutrition and overcrowding, and they and their children lack education. Can anything be done to relieve these conditions of self-destruction which leave the poor no opportunity for improving their mental faculties?

Marshall finds a more comprehensive solution than that of sending good women into the field to develop virtue in the lower classes. The modern industrial system is itself a product of the desired virtues and a few more besides —constant forethought, restless enterprise, deliberate unselfishness—which are all the better for having been freely chosen and not imposed by custom. It is fuelled by free competition, or, to give it what Marshall believed to be a more neutral description, 'economic freedom' (Marshall 1961: I, 9 – 10). The lower classes are to be groomed for and by this system, and to understand this process better, it was necessary to study changes in the 'standard of life', that is, in activities adjusted to wants. This concept involved a development of the ideas

explained by Marshall in his final lecture to the women, that wants should not dictate work (activities), and of the need to find a balance between them. In the *Principles,* Marshall wrote:

> Thus a rise in the standard of life implies an increase of intelligence and energy and self-respect; leading to more care and judgment in expenditure, and to the avoidance of food and drink that gratify the appetite but afford no strength, and of ways of living that are unwholesome physically and morally.
>
> (Marshall 1961:I, 689)

In contrast, the standard of comfort is a term 'that may suggest a mere increase of artificial wants, among which perhaps the grosser wants may predominate' *(ibid.)*.

Many virtues prized by Marshall in both his early and later work are easily recognisable as evangelical. The correct use of time and its use to cultivate the mind by means of certain approved activities (and the severe rejection of others) are easily identifiable, as is the responsibility of parents in the upbringing of their children. Here Marshall identifies the benefits as accruing to society, but is loath to draw the economic consequences of this, since the exercise of parental responsibility forms part of the moral development of the parent. Marshall admits that where competition is free, wealth accumulates rapidly, as does 'misery, crime and all that is bad', but feels this is because the level of culture has not caught up with the new conditions of industrial life (Raffaelli *et al.:* 125). The mentality still is that of slavery not Christianity.[35] The solution is moral, not economic reform.

The lectures to women, their full title being 'Some economic questions directly connected with the welfare of the laborer [sic]', illustrate with great clarity Marshall's overriding interest in improving the moral welfare of the labourer and the subordination of political and economic welfare to this goal. The goal remained paramount throughout his life and he felt that almost anything could be sacrificed to strengthening the moral fibre and cultivating the self-reliance of humankind. He had learned the doctrine of the individual's responsibility for his own salvation through the teachings of evangelicalism to which he had been exposed in his childhood by his father, at school and then at Cambridge, where compulsory chapel attendance several times a week was required of undergraduates. He had proved its worth by succeeding at Cambridge in mathematics through his own efforts and in the face of financial disadvantages, rather than taking the easier path through classics to an Oxford fellowship. At the same time, he had seen how other, richer students wasted opportunities more easily come by.

This heightened sense of individualism even led him to be cautious about formal religion—when he was asked by the women following his lectures about

'sisterhoods', that is, groups of women living together under formal rules and dedicating their lives to charitable works, he condemned them as an 'intense evil' on the grounds that they 'accumulated round them sacred red tape', dwarfed individual character and tended to be dominated by the rule of one unsuitable person or society.[36] Formal religion could thus be a problem since it could involve a slavish devotion to customs and habits.

When in later life Marshall claimed that he believed less in theology but more in religion (Keynes 1985:169), I believe he was referring to the discipline of evangelical moral values which many Victorians who lived beyond the turn of the century felt was hopelessly neglected. He certainly wanted men to feel the need for conversion, to embrace a goal of moral improvement outside themselves, though he no longer regarded this as a route to eternal salvation in some afterlife. He perhaps felt, too, a sentimental yearning for devotional religion, but I believe he rejected the rigidity of mind and regimentation of behaviour which religious practice could involve.

Irrespective of his personal convictions concerning the existence of a divine creator, Marshall clearly did not feel uncomfortable preaching a version of evangelical-Christian morals. This is a trait frequently exhibited by the intellectuals of his generation and sometimes wrongly identified as inconsistency bordering on hypocrisy, since many were known to reject evangelical theology. It is not difficult to understand why they retained the manners and morals of their upbringing, nor that they used a language and imagery with which their listeners were familiar. But I believe there are additional reasons for the persistence of evangelical modes of thought and expression which reflect the nature of their original faith.

Evangelical theology had taught them that to know God was an emotional rather than an intellectual experience—yet Victorian upbringinghad rarely given them access to their emotions. When they finally brought their trained intellects to bear on the question of the existence of God, the thin veil of emotionalism dispersed and they were left with scepticism or total disbelief. Others, who had been brought up on the 'carrot and stick' principle of heaven and hell, found their faith disappearing as parental control over them weakened. Both groups had been reared in a religion which emphasised the role of God on his judgement seat, the doctrine of the atonement by the obedient Son and the (literal) truth of the Bible. The 'new' teaching of the humanity of Christ could therefore be accepted as having worth of its own, irrespective of any theological trappings. For the heaven and hell school, it now became legitimate for heaven to be sought on earth, 'and if heaven was to be sought on earth, then the broad moral implications of Christianity took on a far greater importance than the narrow details of theology—works of charity became more important than belief (Backstrom 1974:19). It was thus that agnostics like Marshall could find so much in common with the theological avant-garde of the church and at times sound so much like them. Even the professionals found him enigmatic on matters of faith.

As the Rev. James Wilson wrote to Mary Paley Marshall when the couple were leaving Bristol for their lengthy stay in Italy where Marshall hoped to rest and regain his health:

> No one has been of late years such a religious teacher to me as [Alfred Marshall]….*Whether he finds help for his life where I do I know not,* and it is a matter of secondary importance. In what is primary he has long been my master.
>
> (29 September 1881, Marshall 1/109, MLA, my emphasis).[37]

Marshall and his generation of intellectuals were criticised by the next as having taken what Lytton Strachey described as 'an easy road to Heaven'. 'What strange people were the married monks of the nineteenth century!…I am alarmed, horrified, impressed—almost over-awed—by such a life' (Strachey's letter to Keynes, 21 October 1924, cited in Skidelsky 1994:182).

One hundred years after Marshall's birth, Jacob Viner remarked that he would not be surprised if Marshall soon became a 'period piece' as a social philosopher (Viner 1982:254). Viner was being polite. Marshall remained a respected economist, but his preaching, based as it was on an evangelical theory of moral progress, was passé. His master-plan for a society of earnest businessmen and workers devoting themselves to honest hard work, self-improvement and the 'noble sources of joy' for themselves and their fellow (English) men, grew increasingly irrelevant as its religious base crumbled, and was totally demolished in the moral degradation of the First World War.

NOTES

1 Wilberforce's 1797 book, *A Practical View of the Prevailing Religious System of Professed Christians, in the Higher and Middle Classes, Contrasted with Real Christianity,* was often considered the 'Bible' of evangelicalism.
2 This description of evangelical dogma is drawn from Bebbington 1989:3 – 17
3 See, for example, Edmund Gosse's reluctance to play with other children because of his 'duty to testify for my Lord, in season and out of season' (Gosse 1986: 117).
4 There are a number of references in biographical literature to sermons on Christ's atonement on the cross for the sins of mankind being preached on Christmas day to the exclusion of the nativity story. Gosse (1986:111) gives a vivid description of the lengths to which his father went to ensure that Christmas was not celebrated in any way in his Plymouth Brethren household of the 1850s.
5 Discussion centres here on the established church since it was overwhelmingly the largest and most influential on the national stage.
6 In his later work (1988:5) Hilton illustrates one of the most important changes of emphasis: 'some evangelicals who had always made a particular fetish of the Cross, began to stress the Christmas message more than that of Easter after about 1840'; cf. note 4 above.

7 The impact of the evangelical movement on Victorian England is something akin to that of the green movement in contemporary Sweden. After a slow start among small bands of enthusiasts, it has grown into a national movement, capable of influencing the policies of all existing political parties as well as public and private industry, and in many cases successfully promoting significant green legislation. Some sections of the green movement are now calling for a *moral* revival, whereby society would make a commitment to environmentally less destructive ways. The 'true believers' remain quite a small group and are split into smaller sects, some of them extremists. Yet the overwhelming majority of Swedes from all walks of life, irrespective of age, sex and political allegiance, use unbleached kitchen paper, return their bottles, batteries and aluminium cans, and recycle their newspapers. In other words, green ways are influencing the outlook of everyone in late twentieth-century Sweden.

8 Dinner parties were just about tolerated because they took place within the confines of that sacred family sanctuary, the home; balls were anathema, '[the] Vicar of Holy Trinity, Cambridge, declaimed against the University's Bachelors' Ball of 1857, contending that a murderer had once been prompted to his reckless crime by the sight of six clergymen at a ball' (Bebbington 1989:131).

9 In his description of a young man, born in 1810, searching for a creed, Young concludes: 'But atmosphere is more than creed, and, whichever way his temperament led him, he found himself at every turn controlled, and animated, by the imponderable pressure of the Evangelical discipline and the almost universal faith in progress' (Young 1989:1).

10 Whickham (1957:127 – 8) describes the 'lower middle-class groups' who flocked to the

> scores of new chapels up and down the town [Sheffield]…natural rallying-points for all who were disposed to take on the decorous habit of chapel-going. It conferred status and confirmed respectability; it provided a centre for like-minded people in an age when organized social facilities were few; it provided the opportunity to 'be someone', and it separated one from the rough lower element in society.

All this and Heaven too!

11 Houghton (1974:186) quotes from Charles Kingsley's *Alton Locke* (1850), the story of a man who began life as a clerk in a grocery store, who

> rose and rose…till…he was owner of a first-rate grocery establishment in the City, and a pleasant villa near Herne Hill, and had a son at King's College [London], preparing for Cambridge and the Church—that being now-a-days the approved method of converting a tradesman's son into a gentleman.

The story is reminiscent of the social advancement planned for Marshall and many of his contemporaries. Marshall's mother was the daughter of a butcher, and

his father, although counting himself a 'gentleman', came from a family which had seen better days. See Coase 1984.

12 Religion seldom made much impact amongst the lowest classes. God and the devil were part of a network of superstitions, all of which could be treated with a certain degree of scepticism. In addition, the overcrowded population of the new industrial towns were not attracted to the established church, whose clergy they suspected were lackeys of a Tory aristocracy, and the work of the methodists amongst the poorest classes was only partially successful. There were also practical difficulties to churchgoing for the poorest members of the population—payment for pews, the requirements of respectable dress, and the collection. Above all lay the difficulty of convincing people living in such degrading conditions of the existence of a loving and merciful God (Houghton 1974:59 – 60). The subject is fully treated in Inglis 1963.

13 A later development was towards a 'sentimentalised version' of God 'less like an Arnoldian schoolmaster and more like Santa Claus' (Hilton 1988: flyleaf). I do not believe Marshall ever accepted such a frivolous version of the Almighty!

14 In the personal search for knowledge of God which preceded the experience of a personal redemption, little importance was attached to the role of priests as authorities in directing the search. Guidance was to be sought from the Holy Spirit, rather than from an external spiritual authority. However, the importance of priests and ministers as persuaders would seem undeniable and they remained figures of authority for those 'weaker' characters and intellects mentioned above—children, women, the lower classes and other races.

15 A further approach involved searching for a 'scientific' proof of the existence of an afterlife and a basis for metaphysics. Henry Sidgwick and his wife Eleanor were closely involved in the Society of Psychical Research. Keynes, unkindly but with some accuracy, wrote of Sidgwick: 'He never did anything but wonder whether Christianity was true and prove that it wasn't and hope that it was'.

16 A Swedish observer, Herbert Tingsten, has described the force of their fear very clearly: 'The gradual withdrawal of God was as terrifying to them [the Victorians] as the atom bomb and other horrors of modern technological warfare are to us' (Tingsten 1972:551).

17 Smith (1967:25) assigns the credit for a long list of late nineteenth-century social reforms to the heroes of his book *The London Heretics*. Ausubel (1960) in his chapter 'Fighting ecclesiastics', provides an antidote.

18 On Victorian dogmatism, see Houghton 1973: ch. 6. Behind this dogmatism lay the anxiety of uncertainty. Henry Sidgwick wrote to H.G.Dakyns in 1866, 'Iwonder if we are coming to an age of general indecisiveness?' (A.S. and E.M.S. 1906:158).

19 Moody and Sankey paid a highly successful visit to Cambridge in 1885, the year Marshall returned to Cambridge as professor. A result of their crusade was the famous missionary expedition of the Cambridge Seven to China, see Pollock 1962.

20 The residue of Simeonites or 'Sims' remaining at St John's College in the 1850s are described in S.Butler 1987:231 *passim*.

21 The repeal meant that it was no longer necessary to subscribe to the thirty-nine articles of the Anglican church to become a college fellow or an MA. The number of Cambridge graduates entering the established church fell from 62 per cent in the period 1800 – 49 to 38 per cent in 1850 – 99 (Jenkins and Caradog Jones 1950:99).

22 Smith (1967:1 – 2) also quotes Ensor (1936) and McKinnon Robertson (1928 – 9) as favouring this date. Green (1964:307) writes:

> If to a remote observer the Universities of the last quarter of the nineteenth century with their compulsory chapel and other religious activities seem strongly impregnated with a religious ethos, it was nonetheless true that the way was now open for further and indeed complete secularisation.

23 Marshall certainly joined his colleagues in studying psychology, but did not follow Sidgwick and others into parapsychology. This sentence written in later life may form part of Marshall's effort to leave behind him a carefully reconstructed intellectual autobiography. An interpretation of his reaction to Mansel could be 'how much there was *in religion* which was worth defending': cf. Marshall telling Keynes how the study of economics 'was a sort of religious work for the sake of the human race' (Keynes to Lydia Lopokova, 16 May 1924, in Hill and Keynes 1989:195).

24 Marshall told James Ward (an erstwhile congregationalist minister who came first in the moral sciences tripos in 1874) that these philosophical studies had got him out of bed at 5 a.m. for three hours' work. They had made him ill and caused his right foot to swell to twice its normal size (Marshall to Ward, 23 September 1904, in Pigou 1925:418). Sidgwick reported a similar affliction in a more light-hearted tone to his mother in February 1866: 'If you hear that I have had an attack of the gout, don't believe it. I have slightly disordered my system with metaphysics and neglect of exercise, and I was obliged to lie up with an inflamed ankle in consequence' (A.S. and E.M.S. 1906:143). The culprit for these ailments was probably not metaphysics, but Cambridge furniture!

25 Butler draws on Richards (1988) who has made a special study of the significance of geometry in nineteenth-century thought.

26 Originally published anonymously, it became clear within months that the author was J.R.Seeley, who became Regius Professor of History at Cambridge after Charles Kingsley's resignation in 1869.

27 For descriptions of the mathematics tripos as championships in memory, handwriting and endurance, see Annan (1984), Strutt (1956), Thomson (1936) and Wilson (1932). Though the records are unclear on this point, Stephen Parkinson of St John's appears at some time to have coached Marshall for the tripos (Pigou 1925: 246). Wilson, a Senior Wrangler, who had the same coach, reports on the concentrated study demanded: 'My tutor would not let me attend Stokes' lecture on the Undulatory Theory and Optics; and was horrified to find out that I was trying to find some fresh advance in Trilinear Co-ordination'. Parkinson wasvery critical of Wilson's reading Mill's *Logic,* Reid's *Philosophy* and Coleridge's *Friend* (Wilson 1932:50 – 1).

28 Keynes based his *Memorial* on information from Mary Paley Marshall (including autobiographical material she had 'rescued from the wastepaper basket'), his father John Neville Keynes, and Cambridge legend, all relating to a period sixty years previously. J.R.Mozley, when asked about the 1860s, admitted: 'The dates…are all mixed up in my head' (J.R.Mozley to J.B.Mayor, 21 April 1904, Add. Ms. c. 104^{66}, HSA). Regarding Marshall's plans to become a molecular physicist, Lord

Rayleigh, Senior Wrangler in Marshall's year, 1865, and later Nobel laureate in physics, was unable to pursue a scientific career in Cambridge at that time because of the absence of laboratory facilities there (the Cavendish Laboratories were not opened until 1874 and remained under-equipped for many years), and of teaching in experimental techniques. Professor Stokes, under whom Marshall hoped to study, was kind and civil but offered Rayleigh no advice or encouragement. A wealthy man, Rayleigh built up laboratory facilities on the family estate, a solution totally beyond Marshall's means (Strutt 1956:37 – 8, 103).

29 Henry Sidgwick had been able to stay at Cambridge after giving up his fellowship at Trinity College, by being made a college lecturer.

30 Marshall probably met Henry Sidgwick for the first time when the latter was examiner at Clifton College, Bristol, in late July 1865 while Marshall was acting as temporary mathematics master there. At Clifton Marshall got to know H.G. Dakyns, who had been a close friend of Sidgwick since their years together at Rugby. Marshall is reported to have admired particularly the practical side of Sidgwick soon after he made his acquaintance (J.R.Mozley to J.B.Mayor, 21 April 1904, Wren Library, Add Ms. c. 104[66], HSA).

31 This emphasis on economic laws as amoral can be found again, for example, in Marshall (1873:109) where providence is called 'nature'; in the *Bee-Hive* articles (1874)—'There are no laws of political economy which can justify our actions'; in the *Economics of Industry* (1879:3, n1); and the *Principles of Economics* (1961: I, 3, para. 4).

32 Alfred West, in a swingeing criticism of Birks' *First Principles of Moral Science: A Course of Lectures Delivered in the University of Cambridge,* described it as 'the worst book on a philosophical subject that we have ever read, and that nothing short of solitary confinement will ever induce us to read such another' (West 1874:452).

33 See Tullberg in Raffaelli *et al.* 1995:47.

34 For example, 'let us remember only that if we squander the "talents committed to our charge" ruin awaits the world' (Raffaelli *et al.* 1995:153).

35 Slavery was an evangelical buzzword prompting an immediate negative reaction. Yet according to Bebbington, it had originally been opposed as a hindrance to mission work rather than on grounds of humanity (see above).

36 Marshall was perhaps thinking of his father's attempts to dominate and control him.

37 James Maurice Wilson, a contemporary of Henry Sidgwick and Senior Wrangler of his year, was a great admirer of Marshall, whom he got to know well when Marshall was Principal of Bristol University College and Wilson was Headmaster of Clifton College. Earlier, he had been a master at Rugby and may well have known Marshall in the 1860s and 1870s through his connections with the Sidgwicks and the Bensons. Should this be the case, it would be further evidence that Marshall was very discreet about his agnosticism if such close friends and colleagues did not know where he stood.

REFERENCES

Manuscript collections in Cambridge: Marshall Archive (Marshall Library of Economics, MLA); Henry Sidgwick Archive (Wren Library, Trinity College, HSA).

A.S. and E.M.S. (Arthur and Eleanor Mildred Sidgwick) (1906) *Henry Sidgwick: A Memoir,* London: Macmillan.

Annan, N. (1951) *Leslie Stephen: The Godless Victorian,* London: MacGibbon and Kee.

——(1984) *Leslie Stephen: The Godless Victorian,* a revised and enlarged version of Annan (1951), London: Weidenfeld and Nicholson.

Ausubel, H. (1960) *In Hard Times: Reformers Among the Late Victorians,* New York: Columbia University Press.

Bebbington, D.W. (1989) *Evangelicalism in Modern Britain: A History from the 1730s to the 1980s,* London: Unwin Hyman.

Butler, R.W. (1990) 'The historical context of the early Marshallian work', paper presented at the Alfred Marshall International Centenary Conference, Florence, December 1990.

Butler, S. (1987) [1903] *The Way of All Flesh,* Harmondsworth: Penguin.

Chasse, J.D. (1984) 'Marshall, the human agent and economic growth: wants and activities revisited', *History of Political Economy,* vol. 16:381 – 404.

Coase, R.H. (1984) 'Alfred Marshall's father and mother', *History of Political Economy,* vol. 16:519 – 27.

Ensor, R.C.K. *England 1870 – 1914,* Oxford: Oxford University Press.

Gosse, E. (1986) [1907] *Father and Son,* Harmondsworth: Penguin.

Green, V.H.H. (1964) *Religion and Oxford and Cambridge,* London: SCM Press.

Hill, P. and Keynes, R. (1989) *Lydia and Maynard: Letters between Lydia Lopokova and John Maynard Keynes,* London: André Deutsch.

Hilton, B. (1985) 'The role of Providence in evangelical social thought', in D.Beales and G.Best (eds) *History, Society and the Churches,* Cambridge: Cambridge University Press, 215 – 33.

——(1988) *The Age of Atonement: The Influence of Evangelicalism on Social and Economic Thought, 1785 – 1865,* Oxford: Clarendon Press.

Hough ton, W.E. (1974) [1957] *The Victorian Frame of Mind,* Clinton MA: Yale University Press.

Inglis, K.S. (1963) *The Churches and the Working Classes in Victorian England,* London: Routledge and Kegan Paul.

Jenkins, H. and Caradog Jones, D. (1950) 'Social class in Cambridge University alumni of the eighteenth and nineteenth centuries', *British Journal of Sociology,* vol. 1:93 – 116.

Keynes, J.M. (1985) [1933] 'Alfred Marshall', in *Essays in Biography,* in D. Moggridge (ed.) *Collected Writings of John Maynard Keynes,* vol. 12, ch. 14, London: Macmillan for the Royal Economic Society.

Marshall, A. (1873) 'The future of the working classes', in A.C.Pigou (ed.) (1925) *Memorials of Alfred Marshall,* London: Macmillan, 101 – 18.

——(1874) 'The laws of political economy' and 'The province of political economy', originally published in *The Bee-Hive,* republished by Harrison, R. (1963) as 'Two early articles of Alfred Marshall', *Economic Journal,* vol. 73:422 – 30.

——(1877) 'The aims and methods of economic study', 9 October 1877. Reported in the *Western Daily Press* and the *Bristol Times and Mirror,* 10 October 1877, and reprinted in J.K.Whitaker (1972) 'Alfred Marshall: the years 1877 to 1885', *History of Political Economy,* vol. 4 (1) spring: 1 – 61.

——(1961) [1920] *Principles of Economics,* ninth (variorum) edition in two volumes, ed. C.W.Guillebaud, London: Macmillan.

Marshall, A. and Marshall, M.P. (1879) *Economics of Industry,* London: Macmillan.

Marshall, M.P. (1947) *What I Remember,* Cambridge: Cambridge University Press.

McKinnon Robertson, J. (1928 – 2) *History of Freethought in the Nineteenth Century,* 2 vols, London: Watts & Co.

Mill, I S. (1873) *Autobiography,* London: Longman.

Norman, E.R. (1976) *Church and Society in England, 1770 – 1970,* Oxford: Oxford University Press.

Oppenheimer, J. (1988) *The Other World: Spiritualism and Psychical Researches in England, 1850 – 1914,* Cambridge: Cambridge University Press.

Parsons, T. (1931) 'Wants and activities in Marshall', *Quarterly Journal of Economics,* vol. 46. References made here are to the reprint in J.C.Wood (ed.) (1982) *Alfred Marshall: Critical Assessments,* vol. 1, London: Croom Helm, 101 – 40.

——(1932) 'Economics and sociology: Marshall in relation to the thought of his time', *Quarterly Journal of Economics,* vol. 46. References made here to the reprint in J.C.Wood (ed.) (1982) *Alfred Marshall: Critical Assessments,* vol. 1, London: Croom Helm, 316 – 47.

Pigou, A.C. (ed.) (1925) *Memorials of Alfred Marshall,* London: Macmillan.

Pollock, C. (1962) [1955] *The Cambridge Seven: A Call to Christian Service,* London: Intervarsity Fellowship.

Raffaelli, T., Biagini, E., and McWilliams Tullberg, R. (eds) (1995) *Alfred Marshall's 'Lectures to Women,* Aldershot: Edward Elgar.

Richards, J. (1988) *Mathematical Visions: the Pursuit of Geometry in Victorian England,* New York: Academic Press.

Rothblatt, S. (1968) *The Revolution of the Dons,* London: Faber.

Scott, W.R. (1924 – 5) 'Alfred Marshall, 1842 – 1924', *Proceedings of the British Academy,* vol. 11:446 – 57.

Skidelsky, R. (1994) *John Maynard Keynes, Volume II: The Economist as Saviour, 1920 – 37,* London: Macmillan.

Smith, W.S. (1967) *The London Heretics, 1870 – 1914,* London: Constable.

Strutt, R.J. (1956) [1924] *John William Strutt: Third Lord Rayleigh,* London: Edward Arnold.

Thomson, J.J. (1936) *Recollections and Reflections,* London: Bell.

Tingsten, H. (1972) [1965] *Victoria and the Victorians* trans. D.Grey and E.Leckström, London: George Allen and Unwin.

Viner, J. (1982) [1941] *Marshall's economics in relation to the man and his times, American Economic Review,* vol. 31, June. Reprinted in J.C.Wood (ed.) (1982) *Alfred Marshall: Critical Assessments,* vol. 1, London: Croom Helm.

Ward, M.A. (1888) *Robert Elsmere,* London: Smith and Elder.

West, A.S. (1874) 'Moral philosophy at Cambridge', *Westminster Review,* vol. XLV: April: 430 – 64.

Wickham, E.R. (1957) *Church and People in an Industrial City,* London: Lutterworth Press.

Wilson, A.T. and Wilson, T.S. (eds) (1932) *James M.Wilson: An Autobiography, 1836 – 1931,* London: Sidgwick and Jackson.

Young, G.M. (1989) [1936, 1953] *Portrait of an Age,* Oxford: Oxford University Press.

5
DISSENTING ECONOMISTS
The late nineteenth-century Indian tradition

Jayati Ghosh

INTRODUCTION

Geoff Harcourt's work—in his writings and teaching—is in the best tradition of dissenting economists. Over the past decades, much of his work has influenced and enriched not just scholarship in 'western' academia, but also Indian discussions which strive to broaden the scope of economic inquiry and make theoretical analysis more relevant to understanding the actual behaviour of economic phenomena. For this reason, and as an appreciative student of Geoff's own generously capacious mode of dissidence, I propose to consider very briefly some of the antecedents to such an approach in the form of contributions made to economic analysis by some nineteenth-century Indian theorists who also, in their own period, questioned the assumptions and conclusions of standard theory.

Why does an economist dissent at all, and choose to think differently from the mainstream tradition? I would argue that there are at least two basic perceptions about economic functioning and analysis which are largely shared by all dissenting economists, and furthermore are at variance with the dominant theoretical framework in economics. The first is the recognition of the inadequacy of the basic tool in the neoclassical economist's kit, methodological individualism. Dissenting economists, for several reasons, see this as an insufficient and often misleading basis for understanding economic phenomena. The Classical economists were all aware that the economy is embedded in society and that economic relations are really no more than an aspect of broader social relations. This in turn implies that it is a mistake to consider the perceived utilities and actions of an isolated individual operating in self-interest—the rational economic man—as truly representative of actual patterns of economic behaviour. The presence of particular institutions and groups, or classes, creates important forms of economic behaviour which are different from those which can be easily handled in standard neoclassical analysis. In terms of applied economicsthat is directed towards guiding policy, this must also lead to an awareness of the influence of social classes in shaping policy and determining its results. Related to this, dissenting economists typically also express

dissatisfaction with the general use of optimisation as the guiding principle. This can be said to derive from the combination of methodological individualism and rational behaviour in a world free of uncertainty, and this combination of assumptions certainly does not justify the universal application of the optimising principle to all 'economic' behaviour. As Sukhamoy Chakravarty—a dissenting Indian economist of more recent vintage—has pointed out (Chakravarty 1992), optimisation is no more than an algorithmic device, and not necessarily a very robust one in many conditions.

Second, there is the importance of time and of history. While neoclassical economists have frequently given formal acceptance to the idea that history matters, there is very little reflection of this in the substantive body of work produced by that tradition. Once the importance of the historical context and historical processes is recognised, it follows that economics as a discipline should necessarily be broadly evolutionary in its approach. This also means that the subject has to absorb the critical importance of time—in the sense of the prevalence of irreversible phenomena—not merely as something which simply exists, but in terms of defining and elaborating the core theoretical concepts. Recent neoclassical models which attempt to capture time and history through analytical devices such as hysteresis, remain limited and in a sense miss the point, because they fail to capture the more inclusive evolutionary perspective.

In the history of Indian economic thought, dissidence—and even outright contestation of standard theory—has always played a key role. This is largely related to the social and political context in which economic theorising has arisen and developed in India. In the late nineteenth century, the study of economics in India was initiated by nationalist writers, not always grounded in the political economy of the time, who were seeking to understand the Indian reality. Their concern with the possibilities of, and constraints upon, the development of a colonial economy definitively coloured their vision, and forced them into dissenting positions simply because their initial assumptions had perforce to be different. The anti-colonialism of their economic conclusions was not simply a knee-jerk patriotic response, but actually emerged from the historical experience of the engagement of the Indian economy with global integration through external trade and colonisation, which was very different from what standard theory predicted. Similarly, the 'initial conditions' of the Indian economy were dramatically different from those which formed the backdrop to the emergence of political economy and later economics in Western Europe. The inapplicability of the standard theories prevalent in England to the Indian context, and their inability to explain patterns of distribution and accumulation in Indian agriculture and industry, made it imperative for Indianeconomists to try and develop their own analyses based upon first principles. Even in the post-Independence era, the specific needs of late industrialisation and the nature of constrained growth in agriculture created the need for more imaginative theorising more suited to specific Indian conditions. The evolution of the

planning process in a mixed developing economy also led to the development of alternative models and analytical methods of more general interest.

This dissent, which was in a sense forced upon the more perceptive Indian economists and others who concerned themselves with economic issues, has actually been extremely beneficial for the tradition generally. In fact, it could be argued that the major insights into economic analysis that have been contributed by Indian economists have ultimately stemmed from this perspective of questioning the existing orthodoxy. These insights have obviously been more relevant for structurally constrained late-industrialising developing economies, but on occasion they have also had more general applicability. In this paper I will consider how dissent shaped the writings of nineteenth-century Indian economists, who have in turn presaged many of the subsequent insights available to economists as late as the final decade of the twentieth century. It is possible to cite at least three areas in which such a tradition of dissent in India has contributed to enriching our understanding of economic processes in general. The first is the area of agriculture and agriculture's relationship to industry and to the broader economy generally. The second relates to public finance and how fiscal policy affects macroeconomics. Finally, there are matters relating external trade, global integration and the position of developing economies in the international economy. Clearly, all these issues cannot really be separated from each other; indeed, one of the characteristics of the more eminent economists in this tradition was their ability to relate processes in one area to another, in an attempt to capture the dynamic behaviour of a constrained developing economy.

THE AGRARIAN QUESTION AND DEVELOPMENT

In a predominantly agricultural economy, the agrarian question must necessarily dominate discussions of economic development. In India, since the nineteenth century, two aspects have been seen as particularly important: the question of land rights and tenures as well as the relationship of the actual tillers of land to the state; and the interaction between agriculture and non-agriculture, in the context of declining or relatively stagnant manufacturing industry.

The economic context of the second half of the nineteenth century was one of a colonial economy experiencing stagnation in total agricultural production and productivity, despite substantial shifts in cropping patterns consequent upon commercialisation and the introduction of large-scale irrigation projects in the limited area of the northwest region. Traditional handicraft industries are estimated to have been experiencing declines in employment, which were not compensated for by increases in employment in the nascent modern manufacturing industries. In addition, there was a spate of famines in various parts of the country particularly towards the end of the nineteenth century.[1] There was an inflationary rise of prices, especially agricultural prices, which began in the 1850s and continued beyond the turn of the century.[2] This was

superimposed on a largely stagnant, low-level economy which was characterised by mass poverty. These circumstances led to a reaction among nationalist Indian writers, against the bulk of writing on agriculture which at the time came essentially from English colonial administrators and foreign academics. The Ricardian theory of rent, which formed the theoretical basis of the utilitarian doctrine espoused by colonial bureaucrats, that in India the state as landlord had rights to the rent of the soil, was rejected by such nationalists as based upon fundamentally different institutional and material conditions. Consequently the level and implications of the state land tax also came up for scrutiny, as well as the relationship between agricultural prosperity—or the lack of it—and industrial development.

Mahadev Govind Ranade (1898)[3] spelt out those assumptions characteristic of the English analyses of the Indian economy which he found unrealistic, and offered an alternative set of assumptions as ingredients for a more relevant theory. In particular, he stressed the relative immobility of both capital and labour across sectors, which would inhibit the operation of the Ricardian mechanism of rent determination. He also posited behavioural assumptions alternative to those attributed to 'rational economic agents'. Thus pursuit of wealth was not the sole aim even in overtly economic activities; custom and the power of the state were more important than market-based competition as economic regulators; and status was more important than contract in determining many economic decisions. In effect, this described a feudal (or at least pre-capitalist) agrarian economy as distinct from a capitalist one. The implications for rent determination were that the land-lease market was relatively inflexible in the short term, rental contracts were determined over a fairly long term, and the level of rents was set by custom which actually imposed relatively high rents. Ranade's ideological position was generally pro-landlord. He argued that the high state land tax operated to lower the expected returns from land ownership, became a disincentive to investment that would increase land productivity, and thus prevented agriculture from modernising. The high prices and buoyancy that nonetheless prevailed in the land market reflected the other factors influencing the price of land—the status associated with it as well as the high personal valuation put on land in an agrarian society.

The solution to all this was seen to lie in systematic and state-guided industrialisation. Taking up from the work of G.V.Joshi (1890), Ranadebrought out the absence of inter-sectoral balance in India and put emphasis on the process of de-industrialisation occurring in the macroeconomy. According to him the latter process, which was closely related to the absence of infant industry protection, was also determined by the miserable economic conditions of the peasantry, which constrained their ability to purchase manufactured goods. Ranade's approach to trade and industrial strategy was heavily influenced by the work of Friedrich List, and he advocated systematic import protection and state aid, without which he felt industrialisation would not gain any significant

momentum. The degree to which such patterns of thought also influenced post-Independence planners in India is well known, and while they may not consciously have been directly influenced by Ranade's work, certainly his position coloured much of the subsequent economic arguments (including those of his protégé Gokhale) during the nationalist movement, and thus created the contours of a general economic consensus by the time of Independence.

R.C.Dutt[4] (1900, 1902) further extended the arguments about land tenures and rental patterns to their connotations for the level of the state land tax. He drew a direct connection between the nature of British imperialism, the land tax imposed by the colonial state and the growing incidence of famines across the country. Dutt rejected the then-prevalent explanations of famines—fluctuations in agricultural output, the excessive pressure of population, the general improvidence of farmers, the depredations of usurers. Instead, he argued that the high state land taxes put undue pressure on the peasantry both directly and indirectly through rental intermediaries, and these high land revenue assessments were integrally related to the problems of widespread rural poverty and recurring famines. His response was to advocate reforms in the colonial administration—land reforms, lower land revenue rates, better economic administration, diminishing military expenditure, increased productive public works as in irrigation, and the like. This of course implied a rather idealistic position on the nature of colonialism, and ignored the economic imperatives of the colonising country. Despite his optimistic reformist bent, Dutt was able to elaborate upon the processes of 'internal drain' from the Indian economy through the combined processes of high taxation and commercialisation of agriculture.[5]

One important analytical link that Dutt made was between the imposition, level and timing of the state land tax and the process of commercialisation of agriculture. He argued that the high and rigidly periodic state land revenue demand in cash at particular times forced the peasantry to sell their produce, often on disadvantageous terms and at unsuitable (i.e. immediately post-harvest) times, simply to pay their taxes. In effect, this presaged the notion of 'forced commercialisation' which has since been studied in the Indian context by a variety of economists such as Krishna Bharadwaj (1985) and Amit Bhaduri (1983). These subsequent analysts have identified a variety of other economic circumstances, in particular indebtedness which can give rise to forced commercialisation, but Dutt's basic characterisation remains useful. The significance of forced commercialisation is that a high degree of magnetisation and selling for the market may actually indicate compulsive involvement in the market, for a variety of economic reasons, without necessarily implying the prevalence of capitalist commodity production for exchange. Also, forced commercialisation, unlike the 'normal' pattern, need have no positive implications for growth or transition to a capitalist industrialisation path. Therefore, those economists who have recognised the widespread prevalence of such perverted patterns of commerce have developed much greater scepticism

about the dynamic, positive role of markets and capitalist behaviour. This has also produced a more sophisticated understanding of the processes of economic transition even among Marxist economists, who (apart from Rosa Luxemburg) otherwise generally tended to ignore or underplay the possibilities of forced commerce.

It is true that Dutt's writings were often highly emotional and floridly descriptive, and his arguments lacked the elegance of presentation that more recent analytical models can provide. Nonetheless, he was able to suggest, in embryonic form, almost all the important insights about famines that are still currently available to economists. Thus, while he did not clearly state anything like the theory of inadequacy of exchange entitlements as an explanation for famines as propounded by Amartya Sen (1981), he was still fully aware of the basic idea that collapses in incomes and purchasing power affected the ability of poor peasants to meet their minimum subsistence requirements and could lead to famines. Further, his discussion of famines was more inclusive than this more recent theory, because he highlighted how long-term changes in local per capita food availability—through the combination of factors like stagnant aggregate production, high taxation, increased exports, changes in cropping patterns, and so on—could reduce average consumption over time and thus weaken the peasantry's ability to withstand any shock, even the expected cyclical fluctuations of weather. This greater proneness to famine as a result of fairly long-term macroeconomic processes, which was hinted at by Dutt, has since been documented for a slightly later period by Utsa Patnaik (1990).

The relationship of agricultural stagnation and poverty to more general macroeconomic processes was also highlighted by G.V.Joshi[6] (1890), who considered the basic structural imbalances of the Indian economy. Joshi's work could be seen as presaging many later structuralist writers concerned with the problems of development and underdevelopment, the more so since he was very conscious of inter-sectoral links. He viewed imbalances within the industrial sector *vis-à-vis* imbalances within agriculture as well as in relation to inter-sectoral imbalances as revealed by the skewed occupational distribution of the population. According to Joshi, there were four main indicators of structural imbalance in the Indian economy in the late nineteenth century. The first was the succession of famines, which were ofgreater severity and frequency than before. The second was the collapse of traditional industries without a concomitant growth of employment in newer modern industries, indicating a process of deindustrialisation. The third was the hike in 'Home charges'—the payment made by the colonial British administration in India to the 'Home government' for the costs incurred by the colonial government in England, in the form of offices, training for bureaucrats, pension payments and the like. Finally, there was the increasing extraction of resources through taxation—what Naoroji was to call the 'internal drain'—from an agricultural population that was already experiencing

substantial material hardship due to the combination of economic stagnation at a low level of productivity and natural calamities.

In this analysis, Joshi made use of several concepts which have a modern resonance. Thus his explanation for agrarian stagnation was that the country was not passing from the extensive to the intensive phase of agriculture. He blamed this absence of transition on the lack of economic incentives for agricultural improvements, owing to an unfavourable system of land relations coexisting with intense pressure on land. The severe competitive pressure for land in turn resulted from the decay of non-agricultural industries and ancillary employment. The absence of alternative income opportunities led to the increasing pressure of displaced population upon agriculture, and further reduced incentives for improved agriculture. These evils were compounded by the sub-division and fragmentation of holdings and the presence of a pyramid of intermediaries in the landholding system. This unhappy combination of forces in turn led to widespread rural underemployment. Joshi was aware of this concept, and even attempted a quantification of the extent of surplus labour in the countryside, based on measures of cropped area per capita and cropped area per agricultural family. It is sobering to reflect on how much of Joshi's analysis remains relevant for India in the late twentieth century.

PUBLIC FINANCE

Throughout the second century of colonial rule in India, the area of public finance was seen as crucial to the state of the economy, both by the government and by nationalist economists, albeit from dramatically different perspectives. Colonial fiscal policy was conservative in the extreme, with an emphasis on balancing the budget, and such balanced budgets were typically hailed in the closing years of the twentieth century as examples of the efficient economic administration provided by the British Raj. The nationalist writer G.V.Joshi, whose work on economic imbalance was discussed above, exploded this particular myth by showing how this excessively prudent fiscal stance actually consisted of patterns of taxation and expenditure which were instrumental in keeping the Indian economy poor and dependent. His twoessays on public finance and public debt (Joshi 1886, 1896) were remarkable in identifying key issues and also in coming to several conclusions which were accepted by mainstream economists in developed countries only a half-century later.

Thus Joshi argued that an economic system may be in equilibrium even though this might involve considerable imbalances. One aspect of such imbalance that Joshi cited (along with those mentioned earlier) was the presence of a large volume of unemployment, as well as widespread economic destitution. The similarity with Keynes' underemployment equilibrium is only too apparent, although Joshi did not develop this proposition in a more theoretical manner. The underemployment equilibrium described by Joshi essentially emerged from the

imbalance between agriculture and industry and the low level of incomes in both sectors, which inhibited inter-sectoral demand for goods. As a result, manufacturing industry was characterised by open unemployment and deindustrialisation, and agriculture by underemployment, or what is today known as disguised unemployment.

Joshi further suggested that the budgetary equilibrium offered nothing more than a summary description of the major macroeconomic processes at work, in India's case those of a dependent economy. His own approach to public finance was not to focus on the simple budgetary equation, but rather to see how consumption, investment and government spending could be related to the structure of the economy. For this purpose, he felt it necessary to study the actual elements and determinants of government spending, in order to know the real purpose of state expenditure. The relation between public expenditures and taxation had also to be uncovered, in specific sectors and as a whole. Joshi was further concerned with the effect of state expenditures and taxation patterns in affecting income distribution and increased inequality. A consideration of these various features led him to argue that in India the colonial government's fiscal policy was both iniquitous (in terms of extracting the most resources from the masses through indirect taxation and land revenue) and anti-growth (because the pattern of expenditure did not emphasise infrastructure building but rather led to outflow of resources and unproductive spending such as military expenditure).

Joshi's method has a surprisingly modern ring to it, and indeed this approach could still usefully be emulated today in several countries where balanced budgets are currently being fetishised. His own consideration of Indian colonial public finance in the light of this approach led him to decry the colonial policy of wasteful and irresponsible spending, especially military expenditures, which were sought to be paid for by increased mass taxation. The burden of increasingly massive, unproductive military expenditure was viewed by Joshi as too high a price for Indians to pay for their own subjection, particularly since this price could be calculated in terms of the foregone alternatives of productive developmental expenditure andsocial welfare spending. In turn, this improvident and reckless expenditure of the colonial government 'determined the level of taxation which bore no normal relation to the tax-paying capacity of the people'. Joshi therefore argued that the real pattern of state expenditure was not 'public' at all, in the sense of operating for the general public benefit, but rather stemmed from the nature and requirements of the colonial state, and was actually inimical to Indian interests.

The management of India's public debt was similarly open to criticism from writers like Joshi and Naoroji. Joshi argued that the way this public debt was handled showed callous indifference both to sound economic principles and to the interests of the mass of the Indian people. There was no attempt on the part of the colonial government to provide for gradual debt redemption or conversion at lower rates of interest. Most of the rupee debt was held by British investors in

England, who received considerably higher yields than those available in the English capital market. It was pointed out by several commentators, including Subramania Iyer (1903), that, given the dearth of capital in India, the fledgling Indian capital market had become oversaturated as a result of the massive increase in ordinary public expenditure, some of which was financed through the issue of government debt instruments. As a result, public loan operations typically benefited foreign investors, both in terms of the yield and the relative safety of the investment, as well as the intermediary foreign banks and speculators, all at the cost of impoverished Indian taxpayers. These assessments provided an insight into the workings of finance capital in a dependent colonial economy.

INTERNAL AND EXTERNAL 'DRAIN'

Many of these different strands of nationalist economic thought came together in the work of Dadabhai Naoroji[7] (1896), whose elaboration of the concepts of internal and external 'drain' from the economy served to provide a unifying framework for dissidence in the Indian context. While Naoroji's own writing was often diffuse and not completely precise, his insights can be formulated and presented together with logical clarity and elegance, as Ganguli (1964, 1977) has shown. Naoroji's formulation of the 'drain theory', and his use of this to explain India's poverty and economic backwardness, ultimately determined the contours of nationalist economic thought in the Independence movement, and was appropriated by leaders such as Gandhi. The idea of an economic drain from India was not new; it had been mentioned from the 1790s onwards, even by Englishmen like Cornwallis—and in the 1830s Rammohan Roy had actually made a rough attempt at quantifying the drain. While Naoroji also essayed a statistical estimation of the external drain, his chief contribution was to link the drain with the behaviour of the balance of payments, and in particular with theexport surplus generated by India throughout the latter part of the nineteenth century. Also, he was able to present this as an explanation of the economic quagmire inhabited by a colonial underdeveloped country.

At the heart of Naoroji's drain theory was the concept of 'unrequited exports'. In an accounting sense India's trade was balanced, but this was only because of the presence of the 'fictitious imports' such as the 'Home charges' and other elements of drain which necessitated substantial financial payments. The need to finance these outflows meant that India's actual recurrent export surplus was not an expression of greater competitive power, but a necessary extraction from the economy. The economic drain (which was composed not only of home charges but also sundry other charges and transfers) was effectively a net unilateral transfer of purchasing power from India to England, and similar in nature to the payment of a tribute. This transferred purchasing power represented an increase in domestic absorption (consumption plus investment) in England, and a

corresponding decline of absorption in India. This idea, that an export surplus can actually be associated with declining incomes and deindustrialisation when there are some limitations on supply, has significant contemporary overtones (Patnaik and Ghosh 1993). Naoroji was critically aware that the important question is not whether the balance of payments balances (for that it must do) but at what level of economic activity it balances, with what composition of transactions, and within what constraints of economic conditions and policies. He also drew attention to the functional link between the budgetary balance and the balance on external account, which is now a mainstay of open economy macroeconomics. The home charges appeared in both external accounts and government financial accounts, which meant that the counterpart of the export surplus was actually the budgetary surplus, with correspondingly depressing effects on the economy.

Naoroji pointed out that the level at which imports and exports balanced in India was relatively low, largely because of the country's economic backwardness. Per capita exports, which consisted essentially of primary commodities, were insignificant in India, even compared to the level reached by some other colonies of the British Empire, in which prolonged capital investments had stimulated domestic production and led to economic diversification such that exports and imports balanced at a high level. The existence of unrequited exports meant that the level of imports was even lower, and this was explained in terms of the low per capita disposable income of the masses, which was itself a product of the drain, and low per capita productivity, which was integrally related to this. Additionally, the inflow of British capital—largely for railway expansion—was not only very small in per capita terms, but it did not really play a constructive role in the Indian context. This was because much of it was actually reinvestment of what had flowed out of India earlier, by way of payment of interest or amortisation of principal.[8] This set up a continuous process of extraction(because of profit repatriation) which exceeded the internal benefits of railway expansion. Thus an initial drain, resulting say from home charges, was the cause of a larger and larger drain over the years.

The phenomenon of internal drain was inseparable from the external drain as exemplified by the export surplus. The vicious circle described by Naoroji had the following elements: the imperative of a unilateral transfer of funds, determined by the colonial government; an increase in sterling indebtedness, resulting whenever exports were inadequate to meet the combined payments for imports and repayment on past debt; a consequent increase in exports to meet the greater repayment demands resulting from the accrual of additional sterling debt; and an increase in taxation in order to extract as much as possible of the real counterpart of the unilateral transfer, in other words to generate an increasing export surplus. This was why the growing export surplus not only coexisted with, but also created, declining per capita availability of basic necessities particularly in rural India, as well as increasing economic hardship faced by the

bulk of the people. This position has been reinforced by modern economic historians like Habib (1974, 1985) in debate with imperialist historians who deny the fact of an economic drain.

This vicious circle was further exacerbated, in Naoroji's opinion, by the decline in the rupee-sterling exchange rate. (The rupee was linked to silver throughout much of this supposed era of 'gold standard' until 1897, and for most of the period from the 1870s silver was depreciating relative to gold.) Although Naoroji did not spell it out very clearly, his discussion certainly anticipated Keynes' later elaboration of the negative terms of trade effect of a unilateral transfer for the transferring country in his (1929) consideration of the transfer problem caused by German reparation payments after the First World War.

The internal mechanics involved in the generation of an export surplus was further explored by G.V.Joshi, who discovered that Indian price trends from 1870 to the late 1890s showed a fall with respect to staple exports (jute, cotton, tea, wheat) and a rise with respect to staple foodgrains of the country (rice, 'inferior' foodgrains, pulses). Joshi argued that this was the result of more land being given over to non-food crops as part of the compulsion to generate increasing exports, even as population pressure increased the local demand for staple foodgrains. The element of compulsion was in turn explained by one aspect of 'forced commerce'—that is, the widespread practice of hypothecation of preferred commercial crops on the basis of advances to farmers by dealers and agents of merchant traders, as well as the increasing importance of plantation agriculture which accounted for a large proportion of agricultural exports.

Much of this discussion continues to be relevant for those countries which primarily export primary commodities, and are forced to make a certain amount of predetermined foreign exchange payments, whether torepay past foreign debt or for other reasons. It also exposes the limitations of 'free trade' practices and the potentially retrogressive effect of global economic integration upon such economies. The arguments of these economists were also informed by a fairly sophisticated perception of the nature of imperialism. Thus Dutt (1898) opined that 'It is an age of imperialism we live in…all over Europe there is an unending struggle for material interest, for conquest, annexations, extension of markets, increase of profits'. It was an assessment similar in broad contours (albeit less analytical) to the arguments of Lenin in the early part of this century. This view of imperialism, of which British colonialism was seen to be one part, was used to understand the behaviour of the colonial regime in India. Thus these early Indian economists never thought of colonial behaviour in simplistic political terms, rather they always related it to contemporaneous economic processes in England and in the world generally.

CONCLUSION

It is evident that the mode of dissent proved to be extremely fruitful for Indian economists writing in the late nineteenth century. This dissidence was not simply a reflection of an urge to disagree with mainstream or established conclusions without adequate cause, but instead sprang from a clear perception of the specific characteristics of the Indian economic reality. This meant that the analysis of these economists provided a definite intellectual framework for understanding internal and external developments of a dependent colonial economy. As has been seen, it also enabled these economists to suggest economic mechanisms and provide the basis for theories that have much wider applicability, anticipate later discussions and give insights that continue to have resonance in the contemporary late twentieth-century world. Dissent was important for these economists, because without a break from established habits of thought it would have been impossible for them to provide such insights.

It is worth noting in this context that the other period of the flowering of Indian economic thought—the post-Independence period when the needs of late industrialisation and economic planning generated some significant contributions to economic theory and policy—was also one when mainstream assumptions and conclusions were continuously questioned. And in this period Indian economists were able to contribute significantly to theories of development planning, to the elaboration of two-sector and multi-sector models which attempted to capture macrodynamic behaviour as well as provide norms for socially optimal investment patterns, and so on.[9] The explicit introduction of the agricultural sector in macroeconomic discussions, and the exploration of inter-sectoral linkages, which was evident in the economic literature of the nineteenth century as discussed above, hasremained a crucial and important contribution by the Indian economic tradition to the general analysis of economic development. This led to the early critiques of Keynes and the Kahn formulation of the multiplier in the context of an agricultural bottleneck, by K.R.V.Rao and A.K.Dasgupta, and more recently has involved formal work on the nature and implications of an agricultural constraint in a developing economy. The political economy of public and private investment behaviour, which has only recently become a focus in mainstream economic discussions in Europe and America, has likewise been an important area of concern for Indian macroeconomists for several decades now.[10]

This fact in itself is of some significance because of the changed attitudes and perceptions of the bulk of economic practitioners in India today. The first phase of global economic integration in the late nineteenth century spawned both ardent supporters and critics, and there was a substantial body of economists and others, in India as elsewhere, who rejected the rosy claims made by the votaries of free trade and imperialist expansion which opened up markets across the world. The current phase of economic globalisation in the last decades of the twentieth century is in some sense broader and deeper than the earlier phase, both because

of the greater volumes of financial capital in international circulation and because of the dominance of large multinational companies that control much of world production and trade. The dominant mainstream paradigm, which bears much similarity with some of the arguments made in favour of the imperialistic 'free trade' of a century earlier, has viewed these international economic developments as positive and beneficial for all parties. In India as well, although there has been a substantial amount of contestation of this position within society and among those who are generally concerned with economic issues, dissent from within the profession has been much more muted and isolated. The tendency within the majority of the profession in India has been the largely uncritical acceptance of the mainstream position.

This is a pity, not only because of the dubious implications this has for the political role of economists, but because new and useful insights into economic processes are unlikely unless standard assumptions can be shaken off. Posterity has always been rather harsh on—or simply forgetful of—intellectual conformists, since they typically contribute little of lasting value. For economists, in India and elsewhere, to make any analytical contribution that will remain relevant, they must rediscover the demanding habits of dissent.

ACKNOWLEDGEMENTS

I am indebted to Claudio Sardoni, Prabhat Patnaik and Arun Ghosh for comments.

NOTES

1 See Morris, Raychaudhuri and others in a special issue of the *Indian Economic and Social History Review,* 1968, for a synoptic discussion and debate on the main economic tendencies in nineteenth-century India.

2 This domestic inflation was strongly related to the adoption in India of the silver exchange standard when the main trading partners were on the gold standard, in a period when silver prices were continuously declining relative to gold. See Atkinson (1909).

3 M.G.Ranade (1842 – 1901), like several other early Indian nationalist thinkers, was a remarkable man who managed to combine proficiency in various different areas with active service in the colonial government. He taught economics in Bombay's Elphinstone College for fourteen years, and subsequently was employed for a very long period (until his death) in the public judicial service. This occupation did not prevent him from continuing his engagement in scholarly study and literary activity, as well as very active participation in the public debates on important policy issues of the time. He was one of the founders and moving spirits of the Poona Sarvajanik Sabha, a nationalist association whose journal published not only many of his own writings but also the writings ofG. V.Joshi and others. Like most of the other economists mentioned here, Ranade was much more significant

historically than even his economic writings would suggest—he was closely involved in, and representative of, the nascent nationalism of educated Indians of the period, who played a substantial role in influencing the course of the modern Independence movement.

4 Romesh Chandra Dutt (1848 – 1909), who was educated at Presidency College in Calcutta, was one of the first Indians to join the colonial Indian Civil service. This was an elite administrative service originally designed to be reserved for Englishmen, and the entry of Dutt along with a handful of other Indians in the late nineteenth century became a contradiction not only for the colonial administration but for Dutt himself During his official career, aside from performing his bureaucratic duties, Dutt wrote mainly historical novels in Bengali, along with one tract on the *Peasantry of Bengal*. He took premature retirement from government service in 1887, after continuously being passed over for promotion by his junior (English) colleagues. Thereafter he embarked on a period of hectic scholarly, literary and political activity in the course of which he published a large number of historical and scholarly works as well as translations of Indian epics. He was known as a fervent land reformer, a campaigner for social causes and a nationalist politician whose views were difficult to ignore because of his practical experience, prodigious scholarly and literary output and his basically reformist (rather than revolutionary) ideological position.

5 The concept of internal drain is elaborated in the discussion on Naoroji later on.

6 G.V.Joshi (1851 – 1911) was a disciple and associate of M.G.Ranade, and with him was influential in moulding the opinions of later important nationalist politicians like Gopal Krishna Gokhale. He was educated at Elphinstone College in Bombay (where he was Ranade's student) and became known as a distinguished scholar. By profession a schoolteacher, Joshi died as Principal of the Poona High School. His writing is marked not only by clarity and a preference for logical arguments over emotional rhetoric, but also for his knowledge of the database and a keen sense of statistical magnitudes. While he is much less wellknown than Ranade, Dutt and Naoroji even among Indian economists today, his work is perhaps even more remarkable than theirs, for the many insights which presage later theoretical and empirical contributions.

7 Dadabhai Naoroji (1825 – 1917) was known during his lifetime as The grand old man of India', and he can safely be described as one of the builders of modern India. His career spanned a very long and eventful period—before the 1857 Rebellion in colonial India he was already professor of mathematics and natural philosophy at Elphinstone College, which spawned an impressive number of nationalist thinkers and eminent persons. He was also subsequently a fairly successful trader, although in the late nineteenth century during the 'Great depression' of prices, Naoroji's business incurred major losses because he attempted to save several of his friends from financial ruin. During the closing years of the Victorian period in England, Naoroji embarked upon a political career, and was actually elected to the House of Commons, as a liberal follower of Gladstone. It was during this time that he sought to elaborate and publicise his position on the 'economic drain' from India. He saw himself, somewhat contradictorily, as a disciplined subject of the British Raj, who wanted to draw attention to the ills consequent upon some aspects of its functioning, rather than as a revolutionary who opposed the basic colonial system.

In his later years, however, he became the doyen of many more radical nationalist activists.

8 The investments by private (usually English) investors in Indian railways were guaranteed by the colonial government, and often these guaranteed profits exceeded the actual returns on such investment. The difference had to be made up by the colonial state exchequer, which of course added to the external drain.

9 The Mahalanobis model of investment planning, developed in the early 1950s (see his collected papers edited by Bose and Mukherjee, 1985) and related contributions by Raj, Sen and Chakravarty are obvious examples. A survey of some of these contributions can be found in Bhagwati and Chakravarty (1969).

10 Patnaik (1995) contains some surveys and discussions of such more recent Indian contributions to macroeconomic analysis for developing countries.

REFERENCES

Bharadwaj, Krishna (1985) A view on commercialisation in Indian agriculture and the development of capitalism', *Journal of Peasant Studies,* vol. 12(4), July.

Chakravarty, Sukhamoy (1992) 'Economics as seen by a dissenting economist', in Philip Arestis and Malcolm Sawyer (eds) *A Biographical Dictionary of Dissenting Economists,* Aldershot: Edward Elgar.

Dutt, R.C. (1898) *England and India,* London: Chatto and Windus.

——(1900) *On Famines and Land Assessments in India.* Reprinted in 1985 by B.R. Publishing Corp. Delhi.

——(1900 – 2) *The Economic History of India, Volumes I and II*, London: K.Paul, Trench, Trübner & Co., Ltd.

Ganguli, B.N. (1964) *Dadabhai Naoroji and the Drain Theory,* Bombay: Asia Publishing House.

——(1977) *Indian Economic Thought: Nineteenth-Century Perspectives,* New Delhi: Tata McGraw Hill.

Habib, Irfan (1974) 'Colonialisation of the Indian economy', *Social Scientist,* vol. 3 (8) March.

——(1985) 'Studying a colonial economy—without perceiving colonialism', *Modern Asian Studies,* vol. 19 (3):355 – 81.

Iyer, G.Subramania (1903) *Some Economic Aspects of British Rule in India,* Madras: Diamond Publishers.

Joshi, G.V. (1886) 'A note on the national debt of British India', *Sarvajanik Sabha Journal,* July.

——(1890) 'The economic situation in India', *Sarvajanik Sabha Journal,* January and October.

Keynes, J.M. (1929) 'The German transfer problem', *Economic Journal,* vol. 39, March: 1 – 7.

Luxemburg, Rosa (1951) *The Accumulation of Capital,* London: Routledge and Kegan Paul.

Naoroji, Dadabhai (1896) *Poverty and UnBritish Rule in India,* reprinted in 1977 by Allied Publishers, Bombay.

Ranade, M.G. (1898) *Essays on Indian Economics,* Bombay: Thacker and Co.

Patnaik P. (ed.) (1995) *Macroeconomics,* Delhi: Oxford University Press.

Patnaik and Ghosh, J. (1991) 'Deindustrialisation with an export surplus', *Economic and Political Weekly,* annual number.

Sen, Amartya K. (1981) *Poverty and Famines: An Essay on Entitlements and Deprivation,* Oxford: Clarendon Press.

6
LIQUIDITY PREFERENCE
Its origin

Pierre Etienne Grégoire

INTRODUCTION

Keynes's liquidity preference theory, which attracted a lot of attention after the publication of the *General Theory,* has recently returned to prominence. The aim of this paper is to provide an appraisal of Keynes's liquidity preference theory and to put it into perspective. The exegesis of how Keynes reached this theory can help us to grasp and understand its originality and importance. Following Keynes's treatment of liquidity preference, we will focus our attention strictly on the demand side even though the supply of money has an important part to play in the determination of the rate of interest.

A comment by Keynes has perhaps given the impression that liquidity preference came very late into the development of the ideas which were to make up the system in the *General Theory:*

> The initial novelty lies in my maintaining that it is not the rate of interest, but the level of incomes which ensures equality between saving and investment. The arguments which lead up to this initial conclusion are independent of my subsequent theory of the rate of interest, and in fact I reached it before I had reached the latter theory. But the result of it was to leave the rate of interest in the air. If the rate of interest is not determined by saving and investment in the same way in which price is determined by supply and demand, how is it determined?…It was only when [attempts in other directions failed] that I hit on what I now think to be the true explanation.
>
> (Keynes 1937:250; see also KCW/XIV:85)

This assertion is true in the same way as an aeroplane becomes commercially useful and effective the moment it carries passengers from one destination to another. But before that day it had to be conceived by engineers, built, testedand so on. The liquidity preference theory also had such a long journey (process), to the examination of which we now proceed.

Trying to pinpoint the specific origins of a particular theory raises numerous problems, especially if economists think that strict continuity characterises the historical development of their discipline. Several stories of the genesis of the liquidity preference theory exist. First, there are those who locate its origins outside the Cambridge school. For example, Edmonson (1970) argues that John Rae used the notion in his monetary theory as early as 1834, though he admits that it did not include the speculative motive. Perhaps following Hicks (1967), Beaugrand (1981) cites Thornton, for whom paper-credit rested on confidence, as a precursor of liquidity preference, regrouping the precautionary motive and the role of anticipations under the demand for bank notes (see Thornton 1978:96 – 7). Keynes himself praised Montesquieu (KCW/VII:xxxiv) and the mercantilists (KCW/VII: 341) for being aware that the rate of interest depended on liquidity preference and the quantity of money. In retrospect, this seems an overly generous reading!

The majority of commentators would perhaps follow Lavington's motto that 'It's all in Marshall, if you'll only take the trouble to dig it out' (Wright 1927: 504) and say that liquidity preference originated within the Cambridge monetary tradition. That Marshall was the leading figure was fully and duly recognised by Keynes in his famous obituary (KCW/X:189). Around 1871, Marshall had set out the essential of his cash balances approach, whereby individuals decide how much money to keep as the result of a balance of advantage between this and alternative forms (see Marshall 1975:166). In this we can find the substance of his later *Money, Credit and Commerce* (1923) where the idea of liquidity preference is presented in the following passage:

> Currency held in the hand yields no income: therefore everyone balances (more or less automatically and instinctively) the benefits, which he would get by enlarging his stock of currency in the hand, against those which he would get by investing some of it…. A large command of resources in the form of currency renders their business easy and smooth, and puts them at an advantage in bargaining; but, on the other hand, it locks up in a barren form resources that might yield an income of gratification if invested.
>
> (Keynes 1923:39 – 40, 45)

In his discussion of the habits of individuals in the *Tract,* Keynes textually followed Marshall:

> Habits are fixed by its estimation of the extra convenience of having more cash in hand as compared with the advantages to be got fromspending the cash or investing it. The point of equilibrium is reached where the estimated advantages of keeping more cash in hand compared with those of spending or investing it about balance.
>
> (KCW/IV:64)

In any case, these quotes only incorporate the transaction and precautionary motives (see KCW/IV:61; KCW/V:200); Marshall did not make operational the speculative motive. Indeed Baumol's transactions demand (1952) is a formalisation of the Marshallian idea which can also be found in Edgeworth (1888), Wicksell (KCW/VII), Hicks (1935), etc. In this paper, however, liquidity preference is regarded as:

1 a theory of money or liquidity;
2 a theory of the demand for money; and
3 a theory of the rate of interest.

In this broader perspective, the liquidity preference theory is a specific positive contribution of Keynes. This does not preclude that he was influenced by others on certain points but, for example, the strong emphasis laid on expectational, subjective and psychic ideas renders Keynes' theory very personal. If this was not the case, why was liquidity preference so badly received within and outside the Cambridge circle?

If everyone agrees with Patinkin (1976:7, 11, 18 – 19) that the *Tract,* the *Treatise* and the *General Theory* constitute a trilogy in Keynes's monetary works, it is perhaps surprising to note that the former book, apart from the equation of exchange, has received little attention on the theoretical or analytical side. We believe that Keynes disseminated in the *Tract* some methodological hints that were to become crucial later on. Although his ideas were not altogether moulded or organised into a global framework, they are essential for understanding the development of this thought.

THE *TRACT*

The third chapter of the *Tract* contains the more theoretical elements of the book. Keynes warned the reader that 'parts of this chapter raise unavoidably matters of much greater difficulty to the layman that the rest of the book' (KCW/IV:61n). Unfortunately, in the Keynesian literature the emphasis was laid on the first part dealing with the quantitative theory, neglecting those dealing with the international aspect of money.

Keynes' explanation of the forward market in exchanges was highly praised by some reviewers. Angell thought it was a 'brilliant and novel analysis' (1925: 273), while for Hawtrey it was 'a topic upon which Mr Keynesspeaks with special authority' (1924:228). A forward contract is nothing more than a spot transaction at a later date fixed on the basis of the spot rate prevailing at the original date. Now what fixes the forward rate, i.e. what determines the sign and the magnitude of the difference between the spot and the future rate? For Keynes, 'the difference between the spot and forward rates is, therefore, precisely and exactly the measure of the preference of the money and exchange market for

holding funds in one international centre rather than in another, *the exchange risk apart'* (KCW/IV:103). The key word here is 'preference'. The subsequent question then was what is it that determines these preferences? He wrote

> the most fundamental cause is to be found in the interest rates obtainable on 'short' money—that is to say, on money lent for or deposited for short periods of time in the money markets of the two centres under comparison. (*ibid:* 103)

Clearly Keynes distinguished between what measures and what determines these preferences, but this should not be understood as an earlier definition (or a theory) of the rate of interest. In that context, Keynes did not deal with the rate of interest; its magnitude is given. Therefore the rate of interest is an *a priori* notion with respect to the spot/forward rates. As the fundamental cause, it pre-exists (temporal relation) the setting of a divergence between the spot and the forward rates. Keynes simply described a causal relationship, specifying that given the rates of interest in two different countries, one detects in the exchange market certain preferences which are mainly reflected or expressed by the divergence between the spot and the forward rates.

Why should we consider this excerpt as the *methodological genesis* of the liquidity preference theory? First, there is the appearance of the term 'preference', a word which conveys a truly subjectivist nature. This expression, though possibly inspired by Marshall's biological approach,[1] is nonetheless Keynes' own and will become crucial later on in his theory of liquidity both in the *Treatise* and the *General Theory*. Second, what lies under the analysis of the forward market is a mode of thinking whereby the impact of a future not yet observed but imagined on the consideration of the present is acknowledged. Third, a close analogy can be drawn between the *Tract* and his later writings. In the *Tract,* preference for holding funds is defined *'the exchange risk apart'* (Keynes' emphasis). In the liquidity preference theory, the rate of interest with which Keynes was dealing was a 'pure' rate of interest.[2] The pure rate is devoid of any allowance for risks (entrepreneur's and lender's) and the like.[3] Even though at first sight it may look a very abstract concept, it is not. For Keynes, following the Cambridge tradition of Marshall[4] and Hawtrey, this concept was related to an institutional framework: the central bank's 'discount rate represents the pure rate of interest' (KCW/XX:535).

The forward market in exchange is by definition time-related, so ideas of uncertainty about the future, and differences of opinions about its course, inevitably come into the picture. The Knight/Keynes distinction between risk and uncertainty, i.e. the incalculable and/or unmeasurable aspect of the latter is made explicit in the *Treatise on Probability* (KCW/VIII:37) and in the *Tract* as well:

> Elsewhere the various uncertainties of financial and political risk… introduce a further element which sometimes quite transcends the factor of relative interest. The possibility of financial trouble or political disturbance, and the quite appreciable probability of a moratorium in the event of any difficulties arising, or of the sudden introduction of exchange regulations which would interfere with the movement of balances out of the country, and even sometimes the contingency of a drastic demonetization—all these factors deter bankers, even when the exchange risk proper is eliminated, from maintaining large floating balances at certain foreign centres. Such risks prevent the business from being based, as it should be, on a mathematical calculation of interest rates; they obliterate by their possible magnitude the small 'turn' which can be earned out of differences between interest rates plus a normal banker's commission; and, being incalculable, they may even deter conservative bankers from doing the business on a substantial scale at any reasonable rate at all.
>
> (KCW/IV:105)

The measurable aspect of risk is perhaps best reflected in the working of the institution, i.e. the use of the forward market as a protective instrument via hedging activities. The example of a reduction in the variability of agricultural prices resulting from 'speculation' in forward markets is widely used. The expression 'mathematical calculation of interest rates' is very helpful to separate uncertainty from probability, a theme previously discussed in the *Treatise on Probability*. With respect to this last point Keynes added that

> he [the banker] has learnt by experience that unforeseen movements of the exchange may involve him in heavy loss; and his prospective profit must be commensurate with the risk he runs. Even if he thinks that the risk is covered actuarially by the prospective profit, a banker cannot afford to run such risks on a large scale.
>
> (KCW/IV:90)

Another very significant factor relating to uncertainty and 'liquidity preference' and concerning the difference between the spot and the forward rates, is the divergence of opinions arising out of the fact that the future course of events is evaluated differently by individuals. The discount on forward dollars will fluctuate between 1 and 2 per cent per annum according as buyers or sellers predominate' (*ibid:* 106). The 'two views' of the stock market in the *Treatise* are here represented by those in the international money market. Two kinds of 'speculators' are distinguished, the 'bear' and the 'bull' (see *ibid:* xxi).

In the *General Theory,* the rate of interest depends on liquidity preference in conjunction with the supply of money, the former reflecting the psychology of the public about the perception of the future. In this respect it is interesting to note

the weight given to the idea that money is a psychological matter, in the preface to the French edition of the *Tract*. The central or pivotal question when dealing with the Keynesian notion of the demand for money is why to hold it besides that needed for payments? The attribute of liquidity, the most essential property of money, makes people 'hold part of their resources in money, because they believe it to be more readily and freely interchangeable than any alternative hoard…if this belief proves wrong, they will not hold money and nothing can make them do so' (KCW/IV:xx).

Thereafter, the notion of liquidity is based on the confidence of individuals; as such, it conveys an idea of non-stability especially when it reaches its most developed form, credit. The role of the economic authorities (governments, central banks and so on) is precisely to preserve this fragile confidence. The instability of the value of money[5] (internal and external) destroys this attitude and modifies what Keynes called the 'social psychology'. For Keynes, human beings facing an uncertain future rely on a sort of law of behaviour, 'the continuous enjoyment of [some thing] engenders an expectation of…[its] renewal' (KCW/IV:18). In other words, people rely on conventions. We assume that the existing state of affairs will continue 'so long as we can rely on its maintenance'. We do so because the status quo is a hard nut to crack (KCW/XIV:125). Convention thus being 'an outcome of the mass psychology of a large number of ignorant individuals' (KCW/VII:154) is inevitably bound to change, and it is this inherent instability that was recognised by Keynes as characteristic of a monetary economy. The notion of convention was already present in the *Tract*. In dealing with the question of the value of the franc, Keynes asked: what has determined and will determine its value? His answer is particularly illuminating,

> First, the quantity, present and *prospective,*[6] of the francs in circulation. Second, the amount of purchasing power which it suits the public to hold in that shape…which depends mainly on the trust ordistrust which the public feel in the prospects of the value of the franc…it is in this second factor, therefore that the crux of the situation lies, namely the attitude of the French public towards their own currency.
>
> (KCW/IV:xviii; emphasis added)

Although the *Tract* does not provide any theory of the rate of interest or a deep analysis of the motives underlying the demand for money, nevertheless we can find in it many pieces of the puzzle around which Keynes was to articulate his liquidity preference theory: the introduction of subjective elements, the importance of uncertainty as distinguished from risk and the way people cope with it through conventions, the fact that individuals differ in their evaluation of the future, the notion of money as a psychological matter. All those notions were to constitute essential parts of the whole Keynesian monetary theory of production.[7]

A careful examination of this early book facilitates comprehension of Keynes' subsequent writings. Many have commented on the intuitive character of Keynes' mind. In this sense, the *Tract* is precisely an intuitive and farsighted book. The importance of the notion of time is latent. This, in our view, constitutes the epistemological problematic which renders Keynes' monetary theory still very much relevant today. Liquidity preference and its accompanying thoughts (anticipation, uncertainty, psychology) are all there waiting to be placed into a more coherent and global framework. Unfortunately, the *Tract* is widely considered the *'parent pauvre* of Keynes' trilogy. Even Kahn held that 'on the analytical side the book was quite traditional' (1975:4). With respect to liquidity preference, we have briefly tried to show that it is not so.

In Chapter 17 of the *General Theory,* there is a particular definition of the money rate of interest as being 'nothing more than the percentage excess of a sum of money contracted for forward delivery, e.g. a year hence, over what we may call the "spot" or cash price of the sum thus contracted for forward delivery' (KCW/VII:222). Bearing in mind Keynes' discussion of what determines the divergence between the spot and forward rates in the foreign exchange market, it is easy to link that analysis to the one presented in the *General Theory*. In the *General Theory,* instead of dealing with money in an international setting via the exchange market, Keynes looked at and examined, in a closed economy, the particularities of money as compared to other commodities. Assets can be classified as being liquid or illiquid; the rate of interest was to be defined as the forward premium to illiquidity. Kregel (1982a, 1982b, 1983), following Davidson (1972, 1980) and Minsky (1975, 1980) in their quest to shift liquidity preference analysis away from the 'traditional' demand for money to the essential properties of interest (as opposed to those of money), has taken that route to put into historicalperspective Chapter 17 of the *General Theory*. For Kregel, the *Tract* exhibits what is now called the 'interest rate parity theorem':

> Forward quotations for the purchase of the currency of the dearer money market tend to be cheaper than spot quotations by a percentage per month equal to the excess of the interest which can be earned in a month in the dearer market over what can be earned in the cheaper.
>
> (KCW/IV:103 – 4)

Kregel translates this passage into the following equality, $(P_f - P_s)/P_s = i_\$ - i_£$=forward discount on dollars. Another way of expressing the same idea is to say

> the return on funds in different financial centres will be brought into equality by means of adjustment in forward rates (and/or interest rate differentials) such that in equilibrium $i_\$(- F_\$) = i_£ + F_£$ where $F_£$ is the forward premium on sterling (and $F_\$$ is zero when the dollar is the standard of comparison and *vice versa*).
>
> (Kregel, 1982a:5)

According to Kregel, this holds in 'ideal conditions' or in equilibrium. Keynes did not lock himself into such a deterministic equilibrium approach. He wrote in the paragraph cited by Kregel 'tend to be', not actually is. Competition (cf. KCW/IV:103) would act so as to bring them into equality, but the time factor must not be overlooked; since history goes on this process may not reach that equilibrium.

An interesting point can be raised in this respect. Before discussing the forward market, Keynes examined the theory of purchasing power parity. He exposed the theory by using a methodological approach which was strictly Ricardian. In its simplest form, that theory stipulates that the rate of exchange between the home and foreign currencies tends in equilibrium to be the ratio between the respective purchasing power of these currencies. Keynes' presentation and analysis implicitly derives from some distinction between the permanence and the transition, e.g. the gravitational method of Smith and Ricardo.[8] There exists a normal exchange rate, a long period equilibrium, around which the actual rate gravitates and converges. 'As time goes on forces will come into play to bring the actual exchange rates and the purchasing power parities nearer together…. This is the point about which the exchanges fluctuate, and at which they must ultimately come to rest' (KCW/IV:72). This point of equilibrium, however, is not fixed because disequilibrium, following different price changes in the two countries, may be brought about not only by a modification of the rate of exchange but also by a movement in the purchasing power parity.

The market or actual rates

> are often more sensitive and more volatile than the purchasing power parities, being subject to speculation, to sudden movements of funds, to seasonal influences, and to *anticipations* of impending changes in purchasing power parity…though also on other occasions they may lag behind.
>
> (KCW/IV:72)

Speculations, volatility and anticipation, three elements on which Keynes will largely expand in his monetary theory, appear within this so-called gravitational framework already quite different from the classical one. Keynes is much more interested in the forces underlying the actual rate, not the natural rate.

Two things must not be forgotten in this discussion of the foreign market. As we have already noted in the *Tract,* the existence of a forward premium or discount on a currency is a measure of the market preferences, the most fundamental cause determining the latter, i.e. the difference in interest rates obtainable on 'short' money. Keynes considered not only the impact of interest

rates differentials because the existence of a fundamental cause does not imply that this is unique. He qualified this by considering very important elements:

1 risk and uncertainty, especially when questions of credit enter the picture (cf. KCW/IV:105);
2 the preponderance of sellers or buyers;
3 'the case, quite frequent in practice, where our assumption of a large and free market breaks down' (KCW/IV:106) and
4 the important case when speculation is active and only in one direction.

Second, there is an inextricable interconnection between the purchasing power parity theory, the forward market in exchange and the quantity theory of money (see KCW/IV:87 – 8). It is in his discussion of the validity of the equation of exchange that Keynes wrote the famous and oft-quoted line on the long run, 'economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is long past the ocean is flat again' (KCW/IV:65).

If Keynes held that real-world problems lie in the transition, why in the two other sections of the same chapter would he contradict himself by being more interested in the long-run implications?

THE *TREATISE*

We have spent much time to extricate the liquidity preference embryo in the *Tract* because it seems to have gone unnoticed in the vast literature on Keynes. The same cannot be said for the *Treatise on Money*. The history of the notion of liquidity preference between the *Treatise* and the *General Theory* has been studied by many, especially since the publication of Keynes' collected writings (see, in particular, Shackle 1967, 1974; Patinkin 1976; Davidson 1972). For our purpose, let us first briefly comment on the 'two views' in the speculative securities market, and then consider Keynes' criticism of the quantity theory.

In deriving his fundamental equations, Keynes had supposed that the price level of new investment goods was given.[9] In his subsequent discussion of the factors upon which this price level depends, Keynes introduced the speculative element of liquidity preference. After deciding what proportion of income should be saved, an individual must decide in what form to hold their 'savings'. In distinguishing these two choices, Keynes brought in an important temporal consideration, because

> the decision as to holding bank deposits or securities relates, not only to the current increment to the wealth of individuals, but also to the whole block of their existing capital. Indeed, since the current increment is but a trifling

> proportion of the block of existing wealth, it is but a minor element in the matter.
>
> (KCW/V:127)

If an individual chooses to hold their wealth in deposits instead of bonds, securities, etc., there is no reason to suppose that this preference will hold in the absolute, independently, for example, of the price of other titles. The preference for deposits (aversion to other titles) depends on the anticipation of the future yield of deposits and securities,[10] and the rate of interest paid on deposits. A change in the relative attraction of deposits and securities can easily happen since the public's attitude changes according to the 'news'. The monetary authority has also an important role via open-market transactions in the creation (destruction) of deposits.

The price level of investment goods thus results from the public sentiment and the behaviour of the banking system:

> a fall in the price level of securities is an indication that the 'bearishness' of the public…has been insufficiently offset by the creation of savings deposits by the banking system—or that the 'bullishness' of the public has been more than offset by the contraction of savings deposits by the banking system.
>
> (KCW/V:128)

There is no definite numerical relationship between the price level of investment goods and the quantity of savings deposits created. In the *Treatise,* hoarding shows a preference for deposits; the ordering of these preferences—i.e. 'the shape of the public's demand curve for savings deposits at different price levels of other securities' (*ibid:* 128)—is what will become the liquidity preference 'curve'.[11] The principal element of variation in the demand for money in the financial sphere comes from the volume of savings deposits. Keynes in the *Treatise* used stock-market language. The psychology of the public with respect to such deposits is polarised around two tendencies or opinions;

> a 'bear'…is one who prefers at the moment to avoid securities and lend cash, and correspondingly a 'bull' is one who prefers to hold securities and borrow cash—the former anticipating that securities will fall in cash value and the latter that they will rise.
>
> (KCW/V:224; see also 239 – 40)

The public's opinion[12] exercises a preponderant influence on the price level of new investment goods. Therefore, in the short term, the price level depends on largely uncontrollable opinions.

With respect to the quantity theory of money, the *Treatise* is a logical prolongation of an argument set forth in the *Tract*. Keynes wanted to get rid of the long-run, straitjacket validity of the quantity theory (see KCW/V: xvii, 120). The analysis in the *Treatise* explicitly distinguished different money deposits implicitly based on different motives for holding money. Contrary to the *Tract,* where the homogeneity of money is reflected in its medium of exchange function, in the *Treatise* the analysis of savings deposits recognises the store of value aspect of money. Referring to his analysis in the *Tract,* Keynes admits that it

> implied that cash deposits are used for nothing except expenditure on current consumption, whereas in fact they are held…for a vast multiplicity of business and personal purposes…its second fault lay in the suggestion that the possible causes of a variation of *k′* were limited to those which can be properly described as a change of habit on the part of the public…. I was applying to the cash deposits, as a whole, conceptions which were only appropriate to the income deposits.
>
> (KCW/V:200)

Thus an interesting feature of the *Treatise* is the much greater emphasis laid on the differences between the behaviour of consumers and business people. Keynes examined the different causes and consequences of shifting money from one category of deposit to another:

> the *Treatise* is, in large part, an analysis of the economy's total need for a stock of money into various sources of demand, and of the mutual struggle of these various component demands each to satisfy itself at the expense of the others.
>
> (Shackle 1967:209 – 10)

This classification of deposits was used to deal with an important question concerning the validity of the quantity theory, i.e. the relation between the volume of deposits and the volume of transactions. Because of the large variability of business deposits (not to mention savings deposits), it is erroneous to suppose that there exists a stable relation between money income and the total of cash deposits (the sum of income and business deposits) (KCW/V:42).

The *Treatise* is first and foremost a theory of fluctuations in the general price level by following an 'income approach'. In the fundamental equations, we do not find the quantity of money in circulation or the velocity of money, two main variables in the quantitative vision. Keynes, however, tried reconciliation with the quantity theory: the unique relation between the quantity of money and the level of prices is a phenomenon which manifests itself only in equilibrium.[13] In the case, for example, of a divergence between the volume of savings and the

cost of new investment, the level of prices may change without any variation in the quantity of money or its velocity,

> In the actual world a change in anything is likely to be accompanied by some change in everything else. But even so the degrees of change in the quantity of money, the velocities of circulation, and the volume of output will not be related in any definite or predictable ratio to the degrees of change in the fundamental price levels. Indeed this is notoriously the case at the acute phases of a credit cycle.
>
> (KCW/V:133)

The importance of the concept of the demand for money *per se* emerged in the analysis of the *Treatise* as an important breakthrough. In this respect, it is important to stress that the quantity theory was *not* strictly speaking a theory of the demand for money.[14] The quantity theory was seen as a theory for the determination of the absolute price level. It raised the question of the demand for money only by detour or ricochet. Keynes, on the other hand, allowed the concept of demand for money to have an existence of its own.

We should also emphasise that Keynes' treatment of the demand for money differs from others because it takes the speculative motive into account. Keynes' theory does not principally refer to the transactions demand. Its essence lies elsewhere. It is around the savings deposits and its'two views' that it takes its existence. What eventually resulted from this perception was a theory of the rate of interest.

In the *General Theory,* the rate of interest was once defined as 'the "price" which equilibrates the desire to hold wealth in the form of cash with the available quantity of cash' (KCW/VII:167). From a methodological point of view it is interesting to compare this definition with a broader point made in the *Treatise:*

> The quantity theory of money has been too often enunciated in a one-sided way, so as to make it appear that the price level depends solely on the volume of money balances created by the bankers. But the price level can be affected just as much by the decisions of the depositors to vary the amounts of real balances which they (the depositors) keep, as by the decisions of the bankers to vary the amounts of money balances which they (the bankers) create.
>
> (KCW/V:204)

The analysis in the *General Theory* seems to be a reduced form of a more general argument set out in the *Treatise;* instead of putting the emphasis on the general price level, Keynes focused his attention on a special price, the rate of interest.

On the history of liquidity preference in connection with the *General Theory,* Moggridge (1973) and Milgate (1977) have provided some very valuable insights; the former with respect to the transition between the *Treatise* and the *General Theory,* the latter by opposing the view that Keynes advanced liquidity preference as a criticism of the classical theory of the rate of interest determined by saving and investment. In this sense, liquidity preference must be seen as a positive and original contribution, a *systemic* concept. Shackle has a strong point in claiming that 'liquidity preference…is the soul of Keynes' theory' (1974:27).

THE *GENERAL THEORY*

Following his analysis in the *Treatise,* Keynes' notion of money in the *General Theory* lies between the restrictive definition of Fisher (1913) and the pure credit model of Wicksell (KCW/VII); money includes deposits, not only cash (KCW/VII: 167 n1).[15] The line between money and debts was drawn subject to the handling of a particular problem.

The speculative motive received more attention because of the causal nexus it provides in linking the industrial and financial spheres. Two conditions are necessary for the existence of liquidity preference. First, uncertainty with respect to future rates. If the future were known, the risk of becoming illiquid or of capital losses would be nil. Second, the fact thatindividuals differ in their assessment of the future.[16] Keynes did not rely on the device of a 'representative' agent. It is the heterogeneity of human beings, not their uniformity which justifies liquidity preference; 'what matters is not the *absolute* level of r but the degree of its divergence from what is considered a fairly safe level of r'(KCW/VII:201). This safe level, which in our view was later wrongly called the normal rate of interest, illustrates the different subjectivity of individuals. Each of them has an expectation of such a safe level; this concept is thus not unique.

How could Keynes discuss the way in which an individual estimates this 'safe level'? He has been often criticised for not providing any behavioural theory of anticipations, but—notwithstanding the use of the notion of 'convention', itself an act of faith to cover a lack of knowledge—there cannot be an *a priori* theory of anticipations since these do not rest on a solid or determinate basis. Shackle (1968:xxiv) has pointed out that 'the *General Theory* has nothing, virtually, to say about how expectations are formed. It cannot say anything, since in its system expectations are the free autonomous variable which governs, and is not governed by, everything else'. Because it is subjective, liquidity preference is prone to change. Keynes more than once stated the psychological character of such a concept. Liquidity preference is not stable or constant; it is a 'potentiality or functional tendency' (KCW/VII:168).[17] The key word is 'potentiality', reflecting the indefinite character of expectations.[18]

Earlier, we mentioned the issue of the origin of Chapter 17 in the *Tract* in conjunction with Sraffa's criticism of Hayek. A few more words on this aspect may be appropriate. First, Keynes' use of Sraffa's concept of own rate of interest is very mechanical, equilibristic in nature and from our point of view not very convincing. I thus totally agree with Harcourt when he says 'I do not really believe it is consistent with his main intuition about the nature of the investment process under capitalist conditions' (1983:82). As for the justification that the money rate of interest rules the roost, Heinsohn and Steiger (1984) have provided an anthropological interrelationship between liquidity and the rate of interest which is much more fruitful.

Finally, some words about the 'finance' motive. This explicit appellation appeared during a debate with Ohlin concerning the determination of the rate of interest. It proceeds from the fact that

> an investment *decision* [Professor Ohlin's investment *ex ante*] may sometimes involve a temporary demand for money before it is carried out, quite distinct from the demand for active balances which will arise as a result of the investment activity whilst it is going on.
>
> (KCW/XIV:207)

To many, this seemed a retreat from Keynes' position. We think that the finance motive is more a change of emphasis than a new concept *per se*. Keynes did not provide a detailed analysis of the finance motive, but nevertheless it is present in his earlier writings. He hinted at it, in the *Treatise* (KCW/V:41), discussing business deposits, and in the *General Theory,* considering the subjective factors that affect the propensity to consume.[19]

Finance is an ex ante concept. It is mainly connected with the volatile or variable character of investment decisions, a central concern in Keynes' theory. In this respect it is interesting to note Keynes' assessment of the debate in a letter to Shaw of April 1938:

> my recent article…did no more than emphasise a little more than formerly… the fact that the finance required by the planning of activity was one of the ways, by no means negligible, in which changes in the level of activity affected the demand for liquid resources, a factor which had always played a prominent part in my theory.
>
> (KCW/XXIX:280 – 1)

Keynes' discussion of the transaction motive was less detailed than in the *Treatise* and, for this, it gave the misleading impression of dealing mainly with current income, whereas as rightly noted by Shackle, 'the transactions motive is an ex ante motive' (1966:264). Keynes' description of the finance motive as a coping stone of his liquidity preference theory was too textually interpreted by

Davidson (1965, 1972), because then he did not have the key information found in the Tilton papers. In the letter to Shaw, Keynes explained what he 'really meant' by that expression:

> I described it as 'the coping stone'…mainly because it seemed to me that it provided a bridge between my way of talking and the way of those who discuss the supply of loans and credits etc. I thought it might help to show that they were simply discussing one of the sources of demand for liquid funds arising out of an increase in activity.
>
> (KCW/XXIX:282)

The real strategic and crucial motive in liquidity preference is the speculative motive.

The purpose of this paper has not been to enter into the critical evaluation of the entire analytical framework of the liquidity preference theory. Instead we have tried to reconstruct an important methodological dimension of Keynes' monetary theory by considering his earlier works. The *Tract* has received particular attention because it provides a vision, an intuition which Keynes later pursued. The basic elements with respect to liquiditypreference are there, waiting to be gathered into a more coherent framework as in the *Treatise* and the *General Theory*. A role for the history of economic thought is precisely to shed some light on the 'cheminement' of an author in the hope that we may learn a way to go beyond.

NOTES

1 When Keynes presented the quantity theory of money in the *Tract* he wrote 'the above method of approach seems less artificial than Professor Fisher's and nearer to the observed facts' (KCW/IV:63n), stressing the point that the Cambridge approach to analysis is based on the individual willingness to hold money. This ex ante motive is governed by a person's desires and decisions, whereas the Yale view is based on a general exchange identity of an ex post nature. For a discussion of the different forms of the quantity theory of money, see Kahn (1984).

2 This is emphasised in a letter to Angell sent in 1938 in which Keynes reiterates that 'liquidity preference, on my definition, relates to the pure rate of interest only' (KCW/XXIX:263).

3 Cf. KCW/VII:145, 208, 221. We shall return to this point later on.

4 Marshall used expressions such as 'mean rate of discount' or 'mean rate of interest'(1923:255).

5 From the outset, Keynes held that 'unemployment, the precarious life of the worker, the disappointment of expectation, the sudden loss of savings, the excessive windfalls to individuals, the speculator, the profiteer—all proceed, in large measure, from the instability of the standard of value' (KCW/IV:xiv).

6 Our emphasis. Some remarks seem appropriate, however. Kaldor (1983) wrote that Keynes made the assumption that the money supply is exogenously given by the monetary authorities. This is a textbook fallacy. How can one write that Keynes' theory follows the tradition of the banking school (Thornton, Tooke, and Hawtrey) and yet say that the supply of money is fixed? The constant money supply in the *General Theory* was a pedagogical device to explain the role of liquidity preference. Harrod, referring to the *Treatise,* argued against this idea (1970:621). Near the end of the famous Pigou article on the quantity theory, there is a little passage which went unnoticed; it concerns the non-independence of the supply and demand of money:

> The demand schedule and the supply schedule for money are not strictly independent of one another…when the element of time is taken into account, a change in the equation of supply may react to alter the equation of demand, and vice versa.
>
> (1917 – 18:62)

By using 'prospective' in such a key extract, Keynes considers the time element. In trying to figure out the future course of the rate of interest, individuals must consider also the actions of the monetary authorities. Will they try to peg interest rates, lend more cash, and raise the money supply? Kaldor, following the long tradition of the banking school [sic], claims that changes in the supply of money are the consequences and not the cause of the changes in prices or incomes. We agree in many cases—but remember how the US government paid for the Vietnam war by printing money instead of boosting taxes!

7 Moreover, the separation between a decision to invest and to save, a key idea in the analysis of the employment/output mechanism, is spelled out, though without further development (cf. KCW/IV:xiv, 9). Most importantly, the whole temporal setting of effective demand based on the entrepreneur's anticipation of recouping their advances *is* in the *Tract:*

> during the lengthy process of production the business world is incurring outgoings in terms of *money*…in the expectation of recouping this outlay by disposing of the product for *money* at a later date…whether he likes it or not, the technique of production under a regime of money contract forces the business world always to carry a big speculative position.
>
> (KCW/IV:33)

8 It is then not surprising to note the similarities of approach between Keynes' analysis of the forward market and Sraffa's criticism of Hayek's *Prices and Production:*

> loans are currently made in the present world in terms of every commodity for which there is a forward market. When a cotton spinner borrows a sum of money for three months and uses the proceeds to purchase spot, a quantity of raw cotton which he

> simultaneously sells three months forward, he is actually 'borrowing cotton' for that period…in equilibrium the spot and forward price coincide, for cotton as for any other commodity; and all the 'natural' or commodity rates are equal to one another, and to the money rate.
>
> (Sraffa 1932:49 – 50)

9 Kahn later criticised this very convincingly; see Keynes KCW/XIII:203 – 7, 212 – 13, 219.

10 Keynes opposes all other forms of wealth to savings deposits. By securities, he considered at the same time loan capital (bonds) and real capital (shares). For the former the link with the rate of interest is known; a rise in their price is equivalent to a fall in the rate of interest and vice-versa. For a share, the same applies within limits (cf. Keynes KCW/VI:174 – 5, 333 – 4).

11 For an idea of such a 'curve' see Keynes KCW/XIII:220 – 1, and also the German or Japanese preface to the *Treatise,* (KCW/V:xxvi).

12 In a market there can be a consensus or differences of opinion. Keynes distinguished four possible types of speculative market (cf. KCW/V:226).

13 There is equilibrium when

> banking habits and practices are unchanged…when the factors of production are fully employed, when the public is neither bullish nor bearish of securities and is maintaining in the form of savings deposits neither more nor less than the 'normal' proportion of its total wealth, and when the volume of saving is equal both to the cost and to the value of new investments.
>
> (KCW/V:132)

Only in this particular equilibrium, $I=I'=S$. In this state, 'the quantity of money available for the industrial circulation *does* rule the situation' (KCW/VI: 4). Thus already in the *Treatise* we find the characterisation of the classical universe as a neutral (real wage) economy which was to become an importantstep towards the *General Theory* (see KCW/XIII; KCW/XXIX). This classical world, where money is only an intermediary link facilitating exchange, is the domain which fostered the quantity theory.

14 Patinkin (1969) has rightly, in our view, used this argument against Friedman in clarifying the Chicago quantity tradition.

15 Keynes stressed in his former book that the difference between demand and time deposits tends to correspond more to differences in banking customs between various clients than to the payment or not of interest. Chick rightly observed: 'criticism of liquidity preference theory on the grounds that speculators would not hold cash when they could hold treasury bills reveals that a combination of careful reading and commonsense did not prevail' (1983:211).

16 Kahn stressed the importance of this aspect earlier on (see KCW/XIV:108).

17 In a letter to Hawtrey, Keynes wrote, 'but of course this is not my theory, for this would mean that I assumed the liquidity function to be constant, which never entered my head' (KCW/XIV:12).

18 '[S]table curves and functions are *allergic* to the real human economic scheme of things'(Shackle 1973:517).

19 Take for example the definition of the motive of enterprise: 'to secure resources to carry out further capital investment without incurring debt or raising further capital on the market' (KCW/VII:108). See also a letter which Keynes sent to Robertson two months before publication of the *General Theory* (KCW/XIV:91).

REFERENCES

References to the Royal Economic Society's edition of *The Collected Writings of John Maynard Keynes* (Keynes 1971 – 89) take the form KCW/volume number: page number.

Angell, J.W. (1925) 'Monetary theory and monetary policy: some recent discussions', *Quarterly Journal of Economics,* 39, February: 267 – 99.

Baumol, W, J. (1952) The transactions demand for cash: an inventory theoretic approach', *Quarterly Journal of Economics,* 56, November: 545 – 56.

Beaugrand, P. (1981) *Henry Thornton: Un Précurseur de John Maynard Keynes,* Paris: Presses Universitaires de France.

Chick, V. (1983) *Macroeconomics after Keynes,* Oxford: Philip Allan.

Davidson, P. (1965) 'Keynes' finance motive', *Oxford Economic Papers,* 17, March: 47 – 65

——(1972) *Money and the Real World,* London: Macmillan.

——(1980) 'The dual-faceted nature of the Keynesian revolution: money and money wages in unemployment and production flow prices', *Journal of Post Keynesian Economics,* 2, 3:291 – 307.

Edgeworth, F.Y. (1888) 'The mathematical theory of banking', *Journal of the Royal Statistical Society,* 51, March: 113 – 27.

Edmonson, N. (1970) 'John Rae and liquidity preference', *History of Political Economy,* 2, 1:432 – 40.

Fisher, I. (1913) *The Purchasing Power of Money,* new and revised edn, New York: Macmillan.

Harcourt, G.C. (1983) 'Keynes' college bursar view on investment: comment on Kregel', in J.A.Kregel (ed.) *Distribution, Effective Demand and International Economic Relations,* London: Macmillan, 81 – 4.

Hawtrey, R.G. (1924) 'Review of *A Tract on Monetary Reform* by J.M.Keynes', *Economic Journal,* 34, June: 227 – 35.

Heinsohn, G. and Steiger, O. (1984) 'Marx and Keynes: private property and money', *Economies et Sociétés,* série Monnaie et Production, 1:37 – 71.

Hicks, J.R. (1935) 'A suggestion for simplifying the theory of money', *Economica,* 2: 1 – 19.

——(1967) *Critical Essays in Monetary Theory,* Oxford: Oxford University Press.

Kahn, R.F. (1975) *On Re-reading Keynes,* London: Oxford University Press.

——(1984) *The Making of Keynes' General Theory,* Cambridge: Cambridge University Press.

Keynes, J.M. (1971 – 89) *The Collected Writings of John Maynard Keynes* (KCW), London: Macmillan.

Kregel, J.A. (1982a) 'Expectations and rationality within a capitalist framework', presented in a course of lectures given at the University of Bremen, June 7 – 11, Cambridge: Marshall Library, mimeo.

——(1982b) 'Money, expectations and relative prices in Keynes' monetary equilibrium', *Economie Appliquée,* 35, 3:449 – 65.

——(1983) 'Effective demand: origins and development of the notion', in J.A. Kregel (ed.) *Distribution, Effective Demand and International Economic Relations,* London: Macmillan, 50 – 68.

Marshall, A. (1923) *Money, Credit and Commerce,* London: Macmillan.

——(1975) *The Early Economic Writings of Alfred Marshall 1867 – 1890,* vol. I, ed. J. K.Whitaker, London: Macmillan.

Milgate, M. (1977) 'Keynes on the "Classical" theory of interest', *Cambridge Journal of Economics,* 1, September: 307 – 15.

Minsky, H.P. (1975) *John Maynard Keynes,* New York: Columbia University Press.

——(1980) 'Money, financial markets and the coherence of a market economy', *Journal of Post Keynesian Economics,* 3, 1:21 – 31.

Moggridge, D.E. (1973) 'From the *Treatise* to the *General Theory:* an exercise in chronology', *History of Political Economy,* 5, 1:72 – 8.

Patinkin, D. (1969) 'The Chicago tradition, the quantity theory and Friedman', *Journal of Money, Credit and Banking,* 1, February: 46 – 70.

——(1976) *Keynes' Monetary Thought,* Durham NC: Duke University Press.

Shackle, G.L.S. (1965) *The Nature of Economic Thought,* Cambridge: Cambridge University Press.

——(1967) *The Years of High Theory,* Cambridge: Cambridge University Press.

——(1968) *Expectations, Investment and Income,* 2nd edn, Oxford: Oxford University Press.

——(1973) 'Keynes and today's establishment in economic theory: a view', *Journal of Economic Literature,* 11, June: 516 – 19.

——(1974) *Keynesian Kaleidics,* Edinburgh: Edinburgh University Press.

Sraffa, P. (1932) 'Dr Hayek on money and capital', *Economic Journal,* 42, March: 42 – 53.

Thornton, H. (1978) *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain,* New York: Kelley.

Wicksell, K. (1898) (KCW/VII) *Interest and Prices (Geldzins und Güterpreise,* trans. R.F.Kahn, London: Macmillan.

W[right], H[arold]. (1927) 'Frederick Lavington', obituary, *Economic Journal,* 37, September: 503 – 5.

7
DENTISTS, DREAMERS AND DEFUNCT ECONOMISTS
Some reflections on English political economy between the wars

Richard Kozul-Wright[1]

INTRODUCTION

It is an interesting feature of post-Keynesian thought that despite its intellectual rotation around a very specific axis, its evolution (and influence) is, to a large extent, due to economists who not only originate from outside Cambridge—which is inevitable—but outside the British Isles—which is rather more intriguing. It has been one of the great contributions of Geoff Harcourt's work to sketch and elucidate those many international connections. However, the study of this intellectual diaspora deserves more attention than it has so far received. This paper does not attempt a fully fledged socio-intellectual history, rather it suggests why post-Keynesian internationalism might be traced back to Keynes' own vision of modern capitalist progress.

Among Keynes' many contributions to the rhetoric of economics, one of his most abiding comes at the very end of *The General Theory:*

> the ideas of economists and political philosophers both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed, the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slave of some defunct economist. Madmen in authority, who hear voices in the air, are distilling the frenzy from some academic scribbler of a few years back. Not, indeed, immediately, but after a certain interval; for in the field of economic and political philosophy there are not many who are influenced by new theories after they are twenty-five or thirty years of age, so that the ideas civil servants and politicians and even agitators apply to current events are not likely to be thenewest. But, sooner or later, it is ideas, not vested interests, which are dangerous for good or evil.
>
> (Keynes 1936:383 – 4)

Most readers who have stopped to reflect on this passage have understood it to be Keynes' final plea to the 'educated bourgeoisie'—among whose ranks he

included himself, and whom he firmly believed held the key to a better future (Skidelsky 1992: ch. 7)—to reject the Classical idea of a self-correcting market economy, and to rethink economic policy along the lines suggested by his own new theory of aggregate demand.

Although Keynes was certainly guilty of this 'rationalist fallacy', this is not an altogether satisfactory interpretation of the close of one of the key economic—and political—texts of the twentieth century. In the first place, it has something of a *post hoc ergo hoc* quality; the golden age of capitalism was ushered in by a new generation of postwar civil servants and politicians committed to a Keynesian agenda and, therefore, introducing policy recommendations latent in *The General Theory* itself. However, although Keynesianism was undoubtedly an important influence on policy-makers in advanced economies after the Second World War, it was not the only one, and in a number of advanced countries certainly not the most important one. But, in addition, this interpretation ignores the stress put in this final passage on the importance of economic and political philosophy and, in the process, abstracts *The General Theory* from Keynes' own efforts to fashion a wider vision of economic and social progress. Closing *The General Theory* with this in mind suggests that Keynes' political economy—its influence on *The General Theory* and beyond—deserves more careful consideration, and that the inter-war policy debates, to which, in part, *The General Theory* was a response, involved more than simply counterposing Keynes to the Classics.

The first two sections of this paper contrast Keynes' political economy with that of Lionel Robbins. The choice of Robbins is deliberate. Whilst Keynes' confrontation with Hayek is perhaps better known today, Lionel Robbins was not only the one British economist offering an alternative vision of modern capitalism, but was also in a position to challenge Keynes in his role as academic and government advisor.[2] Moreover, in light of his comments at the end of *The General Theory,* it is of some interest to note that Keynes had singled out Robbins earlier in the book as one of the few inter-war economists to hold a consistent 'scheme of thought' linking theory and policy (Keynes 1936:20). Despite differences in their visions of capitalist progress and the sometimes hostile nature of their personal exchanges, the third section of this paper will draw attention to some of the similarities between the political economies of Keynes and Robbins; similarities which might, in turn, help explain the international evolution of post-Keynesian thought.

TECHNOCRATIC COLLECTIVISM: THE POLITICAL ECONOMY OF JOHN MAYNARD KEYNES

In the *End of Laissez-Faire,* Keynes identified two dominant ideological tendencies that had shaped economic policies over the previous half-century; 'laissez-faire capitalism' and 'state socialism'. Despite their opposing aims,

Keynes argued that they shared a common intellectual foundation—a political philosophy based on the values of competitive capitalism:

> Nineteenth century State Socialism sprang from Bentham, free competition, etc., and is in some respects a clearer, in some respects a more muddled, version of just the same philosophy as underlies nineteenth century individualism. Both equally laid all their stress on 'freedom', the one negatively to avoid limitations on existing freedom, the other positively to destroy natural or acquired monopolies.
>
> (Keynes 1931:316 – 7)

To this political philosophy was married a distinct conception of the economic problem which emphasised capital accumulation and technical progress as the means to overcome economic scarcity, and on these foundations two very different styles of policy-making evolved; one revolutionary and conflict-based, the other reactionary and unwilling to experiment. But both, Keynes suggested, were deeply pessimistic, and as such posed obstacles to change in the modern world.[3]

Keynes was convinced that any alternative set of policies would have to confront both parts of this vision. His opposition to a narrow emphasis on economic progress and technical change was unequivocal. Indeed, this was his 'central thesis' in *Essays in Persuasion:*

> the Economic Problem, as one may call it for short, the problem of want and poverty and the economic struggle between classes and nations, is nothing but a frightful muddle, a transitory and unnecessary muddle. For the Western world already has the resources and the technique, if we could create the organisation to use them, capable of reducing the Economic Problem, which now absorbs our moral and material energies, to a position of secondary importance.
>
> (Keynes 1931:vii)

and later in a letter to Hayek, he emphatically criticised those who

> ask us to concentrate on economic conditions more exclusively than in any earlier period in the world's history precisely at the momentwhen by their own showing technical achievement is making this sacrifice increasingly unnecessary. This preoccupation with the economic problem is brought to its most intense phase in our evolution when it is becoming less necessary.
>
> (Keynes 1980:386)

The question of Keynes' political philosophy and his approach to the issue of 'freedom' was more ambiguous. Keynes shared with the socialist tradition its

moral commitment to social justice, altruistic impulses and willingness to experiment, and even went so far as to acknowledge the important political cohesion offered by the labour movement.[4] However, the close association between 'acquired monopolies' and class politics and the isolation of the labour movement from policy-making circles made the socialist tradition an unattractive vehicle for the kind of social and economic changes Keynes desired.[5] Intellectually and politically Keynes was both more familiar, and more comfortable, with the liberal idea of personal freedom, and it is this which appears to have played a more significant role in shaping Keynes' own political economy (Keynes 1936:380 – 1).

SKETCHING A MODERN POLITICAL ECONOMY

Although Keynes believed that *laissez-faire* capitalism and state socialism were ideas of a passing era, he was aware that there was no alternative guide to economic policy consistent with the direction and values of the modern age. The urgency of finding such an alternative was already apparent to Keynes when he wrote *The Economic Consequences of the Peace:*

> The forces of the nineteenth century have run their course and are exhausted. The economic motives and ideals of that generation no longer satisfy us: we must find a new way and must suffer again the malaise, and finally the pangs, of a new industrial base.
>
> (Keynes 1984:161)

By the time he wrote *The General Theory,* Keynes had sketched the essential features of an alternative political economy compatible with the evolution of modern society. Keynes' efforts to construct this alternative crystallised around his opposition to the return to the gold standard.[6] Keynes was quick to condemn its immediate impact on British industry. But just as important to Keynes, this decision indicated the persistence among policy-makers of ideas from a bygone age, ideas that continued to pose an institutional obstacle to intelligent policy-making in a new era.

Keynes borrowed the phrase 'era of abundance' from the American institutionalist economist John Commons, to describe the period prior to theFirst World War; an era dominated by capital accumulation and technical progress, economic efficiency and individual liberty, in which the stimulus of competition guaranteed economic and social progress under the leadership of the business classes. For Keynes, this was an essentially 'unnatural' period of economic development. But it was also a necessary one, laying down the conditions through which economic scarcity could be overcome and with it the threats of population growth, diminishing returns and class conflict which had been the predictions of Malthus, Ricardo and Marx respectively (see Keynes 1980:386;

Keynes 1931:361, 369). To the extent it was a self-adjusting system, Keynes believed that the nineteenth-century economy rested on historically specific institutional and psychological conditions. In *The Economic Consequences of the Peace,* Keynes identified two conditions; the free flow of trade and capital associated with the gold standard, and a class compromise which ensured the unequal distribution of wealth generated a process of ongoing accumulation (Keynes 1984: ch. 2). However, his subsequent analysis was focused almost exclusively on the first condition and the pivotal role of British finance capital in organising a global financial system.

In the *Treatise on Money* Keynes referred to the 'dilemma of an international money system' which aimed both to preserve a stable basis for international trade whilst allowing control over domestic economic conditions.[7] But he was using this 'dilemma' to reveal a deeper conflict. By attacking the gold standard Keynes was rejecting the very idea of a self-adjusting economic system in which private and social interest could be made to coincide through market decisions, an idea which he identified with the economic system (and economic thought) of the nineteenth century (see The end of laissez-faire', in Keynes 1931). According to Keynes, the ideological prop beneath Britain's position of 'laissez-faire towards foreign lending' was the assumption that foreign lending would automatically increase the demand for British goods (Keynes 1971:274 – 5). However, with the rise of competing financial and trading centres and given the strength of organised labour in Britain, the 'dilemma' of the gold standard was rapidly shifting Britain's responsibilities inherited from the previous century on to its contemporary labour market problems, the consequences of which, Keynes argued, had not yet been recognised by those responsible for making policy. Exposing the lack of self-adjustment in the contemporary economy was Keynes' first step towards a new political economy and designing a new policy agenda.

However, a necessary element in fashioning an alternative political economy was a clear identification of the real economic problem facing policy-makers in the modern world. We can distinguish at least two possible suggestions in Keynes' writing as to what this problem was. The first sees Keynes at his most romantic, suggesting that the scarcity problem had already been solved by technological progress. This did not exclude temporary adjustment problems—such as technological unemployment—but from this perspective, which received its most eloquent statement in 'Economic possibilities for our grandchildren', economic problems were a diminishing concern rapidly being replaced by the more lasting (and enjoyable) concerns of how best to spend the increasing leisure time which was humankind's destiny.[8]

The second suggestion can be detected in the idea of a mature economy, which plays an important part in much of Keynes' discussion of the British economy but clearly had wider implications[9]. In the *Treatise,* the problems of a mature economy were considered in the context of disequilibrium tendencies associated with an imbalance between saving and investment. Whilst the immediate effect

of such imbalances would be the responsibility of individual entrepreneurs, the banking system through its control of interest rates had an important responsibility in ensuring broader economic stability. Keynes maintained that in a mature economy such as England's, a strong labour movement would add a significant degree of inflexibility to the price structure and ensure that any attempt to orchestrate a monetary deflation to correct economic imbalances would be prolonged. In addition, a declining natural rate of interest, which was characteristic of a mature economy, would produce possible inconsistencies between domestic and international economic goals which had been aggravated with the return to gold: 'In this way adherence to an international standard tends to limit unduly the power of a central bank to deal with its own domestic situation so as to maintain internal stability and the optimum of employment' (Keynes 1971:276). It was from this perspective that Keynes assessed both the need for domestic monetary management as well as its possible limits.

In fact, in these changes shaping the British economy, Keynes saw a more general transition towards an organised form of capitalism, away from the 'economic anarchy' of nineteenth-century regimes and for which a new ideology based on stability and justice was required. These new ideas were described by Keynes, again borrowing from Commons, as fashioning a 'period of stabilisation',[10] providing an alternative to the philosophy of abundance of the earlier age and in which opposition to deflationary pressures implied a new set of responsibilities for policy-makers;

> For the most violent interference with stability and justice to which the nineteenth century submitted in due satisfaction of the philosophy of Abundance, were precisely those which were brought by changes in the price level. But the consequences of these changes particularly when the authorities endeavour to impose them on us in a stronger dose than even the nineteenth century ever swallowed are intolerable to modern ideas and institutions.
>
> (Keynes 1931:337)

What, then, were the modern institutions that Keynes associated with this new era? The first series of institutional changes concerned the reorganisation of modern industry. This took two forms. On the one hand the Marshallian entrepreneur reinvesting profits in the family business was disappearing, replaced by the joint stock company in which 'the general stability and reputation of the institution are more considered by the management than the maximum of profits for the shareholders' (Keynes 1931:314 – 5). On the other hand, the consolidation of organised labour meant that wages were becoming increasingly the result of convention and bargaining over long-term contracts which could guarantee greater stability of labour's position in the economy.[11] Keynes failed to explore in any real detail the managerial or the bargaining

economy. In the process he once again avoided any serious discussion of the question of class compromise which he had introduced in *The Economic Consequences of the Peace*. It would be easy to criticise Keynes for failing to appreciate the particular problems facing industry and for being too quick to condemn its representatives:[12]

> The mishandling of currency and credit by the Bank of England since the war, the stiff-nakedness of the coal owners, the apparently suicidal behaviour of the leaders of Lancashire raises a question of the suitability and adaptability of our business men to the modern age of mingled progress and retrogression.
>
> (Keynes 1981:585)

However, in this context, it is worth recalling that his rather extensive involvement with the reorganisation of the Lancashire cotton industry had left him with little hope of forging the kind of alliance which might revive the supply side of the British economy. Whatever the reasons, Keynes' attitude to industrialists is clearly reflected in his conception of the investment process. For Keynes the determinants of investment were essentially psychological, dominated by 'animal spirits' that mediated the struggle between the commitment to fixed resources and the need for greater liquidity. In the period of transition, monetary conditions were seen by Keynes as the greatest threat to stability.[13]

Not surprisingly, Keynes was concerned to specify the changing nature of the modern financial system. This was a theme across all his major works; the evolution of managed currencies in the *Tract on Monetary Reform,* the extension of the credit economy discussed in the *Treatise on Money,* and the changing relation between finance and industry as reflected in the destabilising influence of the stock market in *The General Theory*.[14] It was in response to these changes that Keynes not only sought to fashion a specific policy agenda but, just as importantly, reached the belief that national responsibilities were paramount over international ones. Once the relation between domestic and international stability had been recognised as acrucial axis of policy there appeared to be a gradual shift in favour of the former. However, it was only with the abandonment of the gold standard that the context for a more stable regime of domestic monetary management emerged. The possibilities of freeing monetary policy from the ideological and institutional obstacles of the nineteenth century, and in a way which would not provoke working class opposition, was now possible. The challenge was to convince those responsible that monetary policy could be consistent with newly evolving responsibilities.[15]

The final feature of organised capitalism was the increasing prominence of public institutions, as reflected not only in the rise of the state and economic management through budgetary decisions, but also in various semi-public institutions. Keynes singled out two such institutions, the Bank of England and

the universities, whose characteristic feature was a commitment to 'the public good' rather than individual gain (Keynes 1971:333 – 5). From this perspective public responsibility essentially involved correcting the mistakes which derived from the old ideology which maintained a hold on civil society (Keynes 1936: 374 – 81). The state as the embodiment of the values of civilised intellectuals (who were increasingly making up its personnel) was a collective technocrat rather than defender of private property, class interests or democratic expression.[16] This interpretation of the public sphere is central to Keynes' policy initiatives and is emphatically stated in the final chapter of *The General Theory* which defended state intervention:

> both as the only practicable means of avoiding the destruction of existing economic forms in their entirety and as the conditions of the successful functioning of individual initiative.
>
> (Keynes 1936:380)

This conclusion was certainly consistent with a concern to convert economists and government officials to a new way of thinking about economic problems, and expanded the possibilities for public responsibility over the direction, though not the detail, of economic development. Furthermore, and despite the rather limited nature of intervention implied by Keynes' notion of public responsibility, it faced strong resistance. One of the most consistent champions of economic freedom was Lionel Robbins, whose contribution has often been overlooked in the attention paid to the conflict between Keynes and Hayek.

THE PERSISTENT CHALLENGE OF LIONEL ROBBINS

Lionel Robbins was born in 1898 into a prosperous nonconformist agricultural family. Following service in the First World War, and undecided as toan appropriate career, Robbins entered the London School of Economics in 1920. After graduating in 1923, he held a number of teaching positions in Oxford and London before being appointed the chair at the LSE in 1929 following the unexpected death of Allyn Young.

It is possible to detect three important influences that shaped Robbins' evolution as an economist. First, a brief association with the Guild Socialists immediately after the War left him with a deep conviction as to the limits of collective action and centralised decision-making.[17] More importantly, as an undergraduate Robbins had been impressed by two liberal critics of Keynes' monetary policy, Gregory and Cannan, who argued that Keynes had—through his resistance to monetary deflation in support of a return to the gold standard—consciously subordinated sound economic advice to political concerns of justice. From this debate Robbins took his commitment to prewar conditions.[18] However, the defensive posture of the British liberal tradition had undermined

the belief in many of its own prescriptions, and in his autobiography Robbins suggests that from the mid-twenties he was searching for a firmer foundations for his own policy discussion. The final influence on Robbins' liberal political economy emerges from his review of Maurice Dobb's *Capitalist Enterprise and Social Progress*. Dobb had located the dynamic element of capitalism in its monopolistic conditions which both supported the innovative impulse of the entrepreneur and provided a position of control over labour in the production process. Whilst Robbins acknowledged the 'very unsatisfactory' theory of production and progress in economic analysis, he was critical of the emphasis Dobb placed on monopoly pressures in the evolution of capitalism. However, in its approach to production, Robbins had identified the Achilles' heel of liberal political economy.[19]

Robbins' search for a more systematic framework within which to develop these influences was completed by the assertiveness and confidence of the Austrian liberal tradition, providing him with the essential elements to develop beyond the 'techniques of thought' of his undergraduate training, into the 'wider sphere' of the relationship between economic thought and political actions. In this respect Austrian economics provided not only an analytical framework but also a wider intellectual vision, or what I have called a political economy. Robbins turned to the Austrians in the late 1920s, and with the arrival of Hayek at the LSE in 1931 their combined influence dominated the direction of the economics faculty.[20] However, Robbins had already been given an important forum for his new ideas on the Committee of Economists.

The Committee had been formed in 1929 by Ramsay MacDonald at the request of Keynes, for whom it represented a technocratic alternative to the larger Economic Advisory Committee, which seemed incapable of producing constructive proposals. Its immediate task was to respond to a number of questions from MacDonald as to the causes and possible remedies of Britain's industrial crisis and particularly the loss of export markets. Keynes had hoped that the recently published *Treatise on Money* would provide the analytical backbone for the Committee's findings. But this proved illusory, and there was continuous disagreement amongst its members (Pigou, Stamp, Henderson and Robbins) as to both the causes and remedies of Britain's economic ills. Despite this disagreement, the Committee did manage to agree to a set of possible policy options, though as Keynes feared these were largely ignored in the confused discussions of the Council (Howson and Winch 1977:76 – 7).

Throughout the Committee's short life, Robbins stubbornly confronted Keynes and not only refused to sign the final report but openly questioned the economic education of his fellow Committee members, suggesting they might benefit from acquaintance with the continental tradition which Robbins had himself recently discovered. Keynes in particular was, according to Robbins, guilty of replacing the intellectual qualities of simplicity and rigour with an obfuscating commitment to aesthetic complexity (Howson and Winch 1977:60).

The resulting disagreements and denunciations further exaggerated the differences on the Committee, with Keynes and Robbins personifying opposing theoretical and ideological positions. The strength of the challenge is fully acknowledged by Keynes in *The General Theory,* where he singles out Robbins as one of the few inter-war economists whose policy recommendations were consistent with their economic theory (Keynes 1936:20). Behind this consistency lay a distinct political economy.

SKETCHING A CONTINENTAL POLITICAL ECONOMY

The basis of Robbins' opposition was subsequently elaborated in two major works; *An Essay on the Nature and Significance of Economic Science,* which was written in 1932, and *The Great Depression,* which appeared two years later. Though subsequently overshadowed by *The General Theory*—and later subject to a well publicised self-criticism (Robbins 1979)—*The Great Depression* was a far from negligible work, as testified by the appreciative, though critical reviews by Henderson and Robertson in *Economica* and the *Economic Journal* respectively.[21] The book was, in fact, an elaboration of Robbins' arguments to the Committee of Economists. As such it combined two basic themes; on the one hand, Robbins insisted that the aim of policy should be a return to prewar conditions and commitments and on the other hand, that excessive state economic management during the postwar period had reinforced the disequilibrium pressures set off by the War. In both respects it represented a powerful and consistent attack on Keynes.

The first of these arguments required Robbins to take a broad historicalperspective on capitalist development in which the century prior to the war was one of uninterrupted economic progress based upon open markets and a strong international monetary system. The War had effected 'a vast series of shifts in the fundamental conditions of demand and supply' by destroying capital and restricting trade, undermining economic flexibility and disrupting growth.

But in addition to the unavoidable influences of the war, the disequilibrium pressures of the 1920s were compounded, according to Robbins, by avoidable cyclical disturbances. In a system governed by the anticipation of profits, he argued, the rate of return on capital dictates the direction of the economy and the relation between natural and money rates of interest will (through its effects on the roundaboutness of production) determine the balance of output between different sectors.[22] Disturbances from either the real or money side can generate cumulative changes building on optimistic anticipations and resulting in over-production during the boom and in eventual and unavoidable collapse, leaving economic imbalances and unrealised expectations. In Robbins' description of the 1920s, the impulse for an unhealthy boom was traced to 'misdirected management on the part of the Federal Reserve authorities' acting in response to Britain's decision to return to the gold standard at prewar parity and more generally

to the 'propaganda for a managed currency' and a stable price level which had accompanied postwar developments and encouraged an unwillingness to see depression follow its natural and appropriate course (Robbins 1934:52 and ch. 5). Monetary disequilibrium was reinforced by disequilibrium in the goods and labour market and identified with institutional changes which disrupted free competition: cartelisation, trade union behaviour and state policy (Robbins 1934: 59 – 60, 83). Not surprisingly, Robbins argued that the abandonment of gold in 1931 and the subsequent increase in tariffs and reflationary measures associated with cheap money policy, simply exacerbated existing disequilibrium pressures.

Underpinning Robbins' discussion was a vision of capitalism whose institutional contours mirrored those of nineteenth-century Britain but whose inspiration was thoroughly Austrian. According to Robbins all societies face a common economic problem—how best to satisfy the insatiable demand of the consumer in an environment of scarcity and uncertainty. In a capitalist economy this problem is resolved through the institutions of the market and private property, which provide the necessary stability, incentives and discipline to allow entrepreneurs to direct resources over which they have ownership rights to the production of goods and services in response to relative price changes.

In such a system, economic responsibility resides with the entrepreneur whose actions in response to profit and loss calculus lead both to technical advance and cost-cutting processes. Moreover, an essential feature of such a system is that such responsibility is thoroughly under decentralised ratherthan centralised control. The success of such a system stems from its flexibility in coordinating the complexity of market behaviours and harnessing competition against the distorting and sclerotic influence of state-induced monopolies (Robbins 1934: 60).

An important aspect of modern society of immediate concern to Robbins' political economy involved the potential clash between (economic) freedom which fostered economic change and (political) democracy which resisted such change. Robbins opposed any interventionist ideology that placed the attainment of economic stability before economic progress, an ideology which legitimated a responsibility of the state over the economy. His two central conclusions were to receive more detailed elaboration a decade later and with greater attention in Hayek's *The Road to Serfdom*. On the one hand, personal and political freedoms are intimately and irreconcilably connected to economic freedoms, while on the other hand, limited state interference with economic freedoms would begin an unstoppable slide to socialism and authoritarianism (Robbins 1934:193).

In this respect it is clear that Robbins already believed that capitalism was being undermined by interventionist ideas which threatened the progressive dimension of markets. His defence of economic freedom was a direct challenge to Keynes:

> It is all very well for the dilettante economists of wealthy universities, their table groaning beneath a sufficiency of the good things of this world, their garages furnished with private means of transport, to say 'Food is cheap enough. Charabancs are vulgar. The railways are admirable'. It is for the millions to whom a slice of bacon more or less, or a bus ride to the sea, still matter, to make the decision.
>
> (Robbins 1934:143)

Robbins' contention that between the market and socialism there was no third way of the kind envisaged by Keynes, assumed an identification of all state intervention and management with planning and socialism. Planning, according to Robbins, had to be total; industrial planning was a contradiction in terms (Robbins 1934:145 – 7). His critique of a planned economy closely followed Von Mises on the impossible calculations confronting central planning. The complex calculations associated with the allocation of scarce resources was beyond the capabilities of any centralised body and could only be processed by decentralised units responding to market signals. Robbins essentially defended the market on second-best terms which opposed any alternative to entrepreneurial responsibility determined by 'technical experts and politicians':[23]

> No doubt capitalism as we know it, encumbered on all sides by interventionism and State-created monopoly, and distorted by thevagaries of mismanaged money, is very far short of the accuracy of competitive adjustment. But with all its inefficiencies in this respect, it seems a much more flexible mechanism than the collectivist alternative.
>
> (Robbins 1934:154)

Robbins' vision did not exclude state responsibilities in a functioning market system; on the contrary, the state would provide critical conditions in support of economic freedom. Rather it was the perceived political and institutional consequences of centralised state action which dictated his approach to policy. Robbins' opposition to public works and his defence of the gold standard both followed this line of reasoning, as did his discussion of agricultural policy in the US (Robbins 1934:129 – 45, 191 – 2; and more generally Robbins 1952)

Having described the institutional configuration of capitalism, the responsibilities of policy-makers and policy options, the final component of Robbins' political economy concerned the nature of change. In this respect Robbins adopted a subjective notion of change, in form (though not in content) not unlike that of Keynes. The problems of modern capitalism were traced to the evolution of the 'socialistic idea', by which Robbins meant the belief that individual actions were incapable in the aggregate of ensuring acceptable social consequences and which therefore required for their realisation a changed institutional framework which could directly represent the social interest. From

this perspective, socialism was not the product of a political movement conditioned by the evolution of modern capitalism with its own distinct ideological and institutional history, but a purely intellectual force, the product of 'detached and isolated thinkers' whose influence could already be traced to policy consequences. Robbins' definition of the socialistic idea was such as to accommodate any notion of collectivism. However, there is little doubt that in this respect Robbins' was responding to the growing influence of Keynes and like-minded theorists. The pressing demand was a reaffirmation of 'individualism'.

This theme was given greater emphasis in *An Essay on the Nature and Significance of Economic Science,* where economic individualism was seen not only as a means to rational action but also as an end in itself. Subsequently, Robbins identified this as a defining feature of classical political economy:

> Classical Political Economists were both individualist as regards ends and (with due reservation) individualist as regards means. For them, an organisation of production based, in the main, on private property and the market, was an essential complement to a system of freedom of choice as regards consumption and provision for the future.
>
> (Robbins 1952:15)

Robbins offered a consistent and for many an attractive political economy. The recovery strategy offered by Robbins, although clearly deflationary in Keynesian eyes—involving cuts in expenditure, a return to the gold standard, freeing trade and reducing costs—was consistent with a political economy centred on progress, individualism and the leading role of the entrepreneur. Moreover, its clear separation of economic and political responsibilities implied that all collective economic outcomes were best fulfilled through individual liberty (Robbins 1971:161). This was a powerful picture, and as recent interpretations of the Treasury view suggest, one that was far from marginal to the development of policy in the inter-war period.[24]

THE LIMITS OF ENGLISH POLITICAL ECONOMY

In an important essay, Peter Hall has suggested that the influence of Keynesian ideas cannot be divorced from particular national circumstances (Hall 1989: ch. 14). But the importance of the national context also extends to the evolution of Keynes' own vision of modern capitalism; this not only included the particular problems Keynes confronted, for example on the supply side of the British economy, or the semi-independent role of the economic bureaucracy in Britain, but also the need to challenge the strong alternative vision offered by Robbins. Moreover, as Joseph Schumpeter recognised, there was a distinct national bond

among many English economists of the inter-war era (Schumpeter 1949) which in some important respects transcended these differences in political economy.

In the first place, both Keynes and Robbins saw the economic problem as one of pleasing the consumer. For Robbins, this was firmly grounded in his optimistic assessment of market exchanges between individuals freely pursuing their own particular interest, and later reflected in his lasting definition of economic science. But for Keynes, too, consumption was the only truly meaningful objective of economic activity. Consequently, for both writers, the economy was, in essence, a sphere of harmonious interests, albeit one that could temporarily be disturbed. Certainly, for Keynes the 'bank-balance' mentality of the majority would still on occasion have to be controlled through the technocratic collectivism of an enlightened bureaucracy and other appropriately staffed public bodies acting for the social good. While for Robbins, faith in individual choice still gave greater weight to the competitive forces of the market. But it is not altogether surprising that following his experience as an administrator of the wartime economy, Robbins' differences with Keynes began to focus around the best way to realise an optimal collective outcome out of the present vicissitudes (Schumpeter 1949:307); a convergence of views which was to receive a fuller intellectual expression in much conventional postwar Keynesian thinking.

Related to this common conception of the economic problem, a second strand woven into the works of Keynes and Robbins was the pre-eminence both writers gave to defending individual freedom. The notion of selfish and egotistical 'man' had dominated economics since Smith, and it found full expression in Robbins' idea that the rational entrepreneur must be given unrestricted freedom to make full use of his assets in the best way he saw fit. But an equally old tradition conceived of 'man' as essentially altruistic and public spirited, motivated by concerns of social justice. Keynes combined both these ideas in a bifurcated world of individual (economic) freedom in civil society and altruism and experiment in the public sphere, of which the state was the highest though not the only guarantor (Keynes 1936:374). In this respect, Keynes' individualism shared much in common with the Bloomsbury Group, with which he was quite intimately associated. In an important essay, Raymond Williams has argued that the importance of Bloombsury lay in their cultural challenge to nineteenth-century values (broadly conceived to include moral, theoretical and ideological concerns) and their elaboration of an alternative set of values which appealed to the supreme value of the civilised individual, whose multiplication was seen as the best hope of progress in modern society (Williams 1989). The parallel with Keynes' political economy is striking.

A third common strand between Keynes and Robbins was their belief in the power of ideas over political interests. Keynes expressed this belief forcefully in the final passages of *The General Theory*. But a remarkably similar expression of this position closes *The Great Depression,* where Robbins complains of those

'detached and isolated thinkers' whose creation of the socialistic idea was the great distortion of his age:

> men of intellect, with powers of reason and persuasion, have conceived the socialistic idea and gradually persuaded their fellows… In the short run it is true, ideas are unimportant and ineffective but in the long run they can rule the world.
>
> (Robbins 1934:200)

Perhaps not surprisingly, and like Keynes, Robbins put his faith in the patience of educated minds.

This takes us to a final point in common between Keynes and Robbins; their shared notions of the political realm. The striking connection between them in this respect was their shared failure to focus on group values, and particularly those related to the production process (Hirsch 1978). Despite Robbins' search for a convincing response to Dobb and Keynes' experience with the reorganisation of the textile industry, for both writers production relations had a shadowy existence. Not surprisingly, for both Keynes and Robbins the organised labour movement was excluded from responsibilities over or involvement in the policy process. Also in this context, it should notbe forgotten that an important feature of the macroeconomic adjustments outlined by Keynes was the avoidance of any bargaining between powerful groups with conflicting interests, in favour of an imposed collective outcome;[25] this was in line with the tradition of public responsibility common to both Keynes and Robbins (see Hirsch 1978). As Williams (1989) has suggested in his discussion of the Bloomsbury Group, this approach can, in part, be explained by the growing need for a more professional and educated administrative stratum to manage the increasing responsibilities of the state, a demand which lay behind reforms to the British universities, including Cambridge and the London School of Economics. This political context certainly helped foster a position which idealised the state as a defender of individual freedoms and responsibilities, and also prejudiced both Keynes and Robbins against more radical reforms to the state as a prerequisite for their respective policy agendas.[26]

A corollary of this technocratic collectivism was its limited policy dialogue, both in terms of the issues discussed and the economic voices included in the policy-making forum. As Schumpeter (1949) noted, few English economists have ever taken socialism seriously, and this was particularly true of both Robbins and Keynes. Moreover, neither showed more than a passing interest in democratic values. Indeed, if any attention was paid to democracy, it was seen by both as a hindrance to effective policy making. But just as significantly, to the future of the British economy the question of industrial policy was little discussed by both writers.

CONCLUSION

Keynes believed that the early twentieth century was both a period of profound socioeconomic transition and one which was lacking an intellectual vanguard. During the inter-war period conceptions about the nature of society and the economy, the role of government, political ideals, as well as past policy experience were undergoing dramatic reassessment. Political economy was in flux. These challenges are not captured in the struggle between Keynes and the Classics, and I have suggested that the famous passages at the end of *The General Theory* point in a different direction. In particular, I have argued that the differences between Keynes and Lionel Robbins better capture these struggles. However, and whether or not Keynes really freed himself from his own nineteenth-century intellectual legacies, his conflict with Lionel Robbins does highlight the importance of a particular national context to the evolution of Keynes' own thinking. In this respect, his vision of modern capitalism was seriously compromised by the absence of a clear theory of the state and of a clear understanding of the links between politics and the sphere of production. Perhaps not surprisingly, it was the marriage of Keynesian policies to other political philosophies whichproved more conducive to a progressive political economy after the Second World War, and which, in part, helped propel post-Keynesianism on its international trajectory.

ACKNOWLEDGEMENTS

The author benefited from comments by Rugerio Studart and Jan Kregel on an earlier draft of this paper.

NOTES

1 The opinions expressed are those of the author and do not reflect those of the United Nations Secretariat.
2 In this respect the growing influence of academic advisors to government provides an important backdrop for the discussion of political economy in the inter-war period. See Winch 1969.
3 See, in particular, *Economic Possibilities for Our Grandchildren,* in Keynes 1931.
4 See 'Liberalism and labour', in Keynes 1931; and The dilemma of modern socialism' (Keynes 1982:33 – 3). Skidelsky (1988) has argued against the importance of 'equality' in Keynes' thought and suggests that the idea of social justice was a youthful Burkean legacy. However, there was more of a genuine tension in these aspects of Keynes' thinking which was embedded in his political economy.
5 Keynes' attitude was undoubtedly coloured by the peculiarities of the British Labour Party. However, there was never on his part any serious attempt to understand the intellectual diversity of the socialist tradition and particularly European social democracy.

6 Despite its subsequent role in Keynesian thinking, unemployment was not Keynes' immediate focus, though its relation with the gold standard was never far away. See the important discussion in Hancock 1960, for a critical appraisal of inter-war British economists on the unemployment question.
7 See Keynes 1971:270 – 2; 1931:208, 262. Keynes first made this argument in *A Tract on Monetary Reform* though very much within the traditional quantity theory framework.
8 Keynes 1931. This romantic streak was not, it should be emphasised, peculiar to Keynes.
9 The much discussed stagnation thesis in Keynes can be seen as a pessimistic version of the mature economy thesis.
10 Regardless of whether the effect of a savings-investment imbalance was felt through changes in prices or output. Discussion in the *Treatise* was largely confined to price changes.
11 On the increasingly rigid nature of wages see 'Industry and liberalism', in Keynes 1981; also Winch 1989.
12 See his comments on trade unions in 'Essays in persuasion', Keynes 1931.
13 On Keynes' experience with the cotton industry see Wright 1989: ch. 2.
14 As Moore (1988) has carefully discussed, there remained considerable ambiguity as to the 'exogenous' or 'endogenous' nature of money throughout Keynes' writing and consequently the influence of the central bank in determining aggregate performance.
15 It is important to note that monetary policy was seen by Keynes as a means of avoiding economic instability. If the economy was in an extreme disequilibrium position already, other measures (such as public works or the 'socialisation of investment') might have to be taken. However, his commitment to domestic monetary policy was a consistent theme throughout Keynes' writing; see Booth 1989; Howson 1975; De Brunhoff 1989.
16 Keynes' relation to the Bloomsbury faction is critical in this respect; see below. Keynes' justifications for forming the EAC was its freedom from 'party attitudes' and conflicting economic interests (see Howson and Winch 1977). Thomas (1984: 377 – 8) has suggested that part of Keynes' failure (and that of academic economists more generally) to influence policy in the inter-war period might, in part, stem from the limited impact of the Committee of Economists.
17 See Robbins 1971. Like Keynes, Robbins never joined any political party.
18 See Hancock 1960:306; Winch 1969:106. One of Robbins' first articles was an 'unprovoked attack' on Keynes along similar lines to Gregory and Cannan (see O'Brien 1988:106).
19 See Robbins 1926 and the response by Dobb in the same issue of *Economica*. Dobb had been a Ph.D. student at the LSE and the book grew out of his thesis.
20 Pimlott (1986:161) writes that 'in the intensely competitive atmosphere of the Hayek-Robbins weekly seminar, alternative views had little place'. Of course, this was the period in which the LSE rose to challenge Cambridge; Hicks, Kaldor, Durbin, Lerner, Coase and Lewis were all students at the school during these years. Many were subsequently to become supporters of Keynesian economics.
21 Both reviews took issue with Robbins' deflationary prescription and his unrelenting anti-statism.
22 See Robbins' subsequent comments on Wicksell in Robbins 1970.

23 Robbins 1934:154 – 5. Here was the missing theory of production and progress identified in his debate with Dobb.
24 I have suggested elsewhere its importance for the formation of postwar British macroeconomics (see Wright 1989).
25 'Socialising' investment, by which Keynes meant maintaining a low rate of interest, with unspecified additional intervention if this failed to achieve its desired effect, is a clear example. See also Hirsch 1978.
26 In this respect, searching in Keynes for a theory of the state is almost certain to be a disappointing one (see Skidelsky 1988; De Brunhoff 1989).

REFERENCES

Booth, A. (1986) 'Simple Keynesianism and Whitehall, 1936 – 47', *Economy and Society,* 15, 1 – 22.
——(1989) *British Economic Policy, 1931 – 49,* London: Harvester.
De Brunhoff, S. (1989) 'The Keynesian critique of laissez-faire', in A.Barre (ed.) *Keynesian Economic Policies,* London: Macmillan.
Hall: (1989) 'The politics of Keynesian ideas', in P.Hall (ed.) *The Political Power of Economic Ideas,* Princeton NJ: Princeton University Press.
Hancock, I (1960) 'Unemployment and economists in the 1920s', *Economica,* 27.
Hirsch, F. (1978) *The Social Limits to Growth,* London: Routledge and Kegan Paul.
Howson, S. (1975) *Domestic Monetary Management in Britain, 1919 – 38,* Cambridge: Cambridge University Press.
Howson, S. and Winch, D. (1977) *The Economic Advisory Council: A Study of Economic Advice in Depression and Recovery*, Cambridge: Cambridge University Press.
Keynes, J.M. (1931) *Essays in Persuasion,* London: Macmillan.
——(1936) *The General Theory of Employment, Interest and Money,* London: Macmillan.
——(1971) *A Treatise on Money,* London: Macmillan.
——(1980) *Collected Works, Volume XVIII: Activities 1922 – 31,* London: Macmillan.
——(1981) *Collected Works, Volume XIX: Activities 1922 – 9,* London: Macmillan.
——(1982) *Collected Works, Volume XXVIII: Social, Political and Literary Writing,* London: Macmillan.
——(1984) *The Economic Consequences of the Peace,* London: Macmillan.
Moore, B. (1988) *Horizontalists and Verticalists,* Cambridge: Cambridge University Press.
O'Brien, D.P (1988) 'Lionel Charles Robbins, 1898 – 1984', *Economic Journal,* March.
Pimlott, B. (1986) *Hugh Dalton,* London: Macmillan.
Robbins, L. (1926) The dynamics of capitalism', *Economica,* June.
——(1934) *The Great Depression,* London: Macmillan.
——(1952) *The Theory of Economic Policy,* London: Macmillan.
——(1970) *The Evolution of Modern Economic Theory,* London: Macmillan.
——(1971) *Autobiography of an Economist,* London: Macmillan.
——(1979) *Against Inflation,* London: Macmillan.
Schott, K. (1985) 'The rise of Keynesian economics, Britain 1940 – 64', *Economy and Society,* August.
Schumpeter, J. (1949) 'English economists and the state-managed economy', *Journal of Political Economy,* October.

Skidelsky, R. (1988) 'Keynes' political legacy', in O.Hamouda and J.Smithin (eds) *Keynes and Public Policy after Fifty Years,* Aldershot: Edward Elgar.

——(1992) *John Maynard Keynes: The Economist as Saviour, 1920 – 37,* London: Macmillan.

Thomas, M. (1984) 'Discussion', *The Journal of Economic History,* XLIV, June.

Williams, R. (1989) The Bloomsbury Group', in *Essays in Materialism and Culture,* London: New Left Books.

Winch, D. (1969) *Economics and Policy: A Historical Study,* London: Hodder and Stoughton.

——(1989) 'Keynes, Keynesianism, and state intervention', in P.Hall (ed.) *The Political Power of Economic Ideas,* Princeton NJ: Princeton University Press.

Wright, R. (1989) Three essays in comparative institutional economics: Britain and Sweden since 1918', Ph.D. thesis, University of Cambridge.

8
KEYNES'S SOCIALISM
Conception, strategy and espousal

Rod O'Donnell[1]

This chapter addresses the way in which Keynes conceptualised his own political thought. The issue lies at the heart of Keynes's politics but has not received the analytical attention it deserves. It is relevant to understanding his economics, politics and social vision, and also to modern debates over the future of socialism.

My primary thesis is that Keynes, in his writings, advocated a particular kind of socialism which, to use his own term, may be called *liberal socialism*. This was a new form of socialism aimed at fusing the best elements in both liberalism and socialism, and discarding the worst. Associated with his advocacy were critiques of existing forms of liberalism, particularly *laissez-faire* liberalism, and of existing forms of socialism, particularly Marxist socialism and State socialism.

The thesis forms part of a larger argument comprising three propositions:

1 Keynes envisaged and espoused a particular form of socialism.
2 He advanced an accompanying general strategy for its achievement.
3 His version of socialism constitutes a legitimate form of socialism.

This paper, however, is dedicated to the first and second propositions only. The third proposition is far too large a task to be undertaken here. It requires, firstly, the provision of a satisfactory general definition of socialism, the absence of serious attempts at which is one of the great internal weaknesses of socialist thought. Secondly, it requires an argument showing that Keynes's variant falls within the scope of such a definition (at least to the same extent as other recognised members of the socialist family). Pending demonstration of the proposition that Keynes's liberal socialism satisfies relevant criteria for admission to the socialist family, the present paper confines itself to the nature of Keynes's version of socialism and his steady espousal of this version over many years.

A subsidiary thesis is that Keynes's form of socialism continues to have value and instructiveness for the political left of modern times. Recentevents have increased, rather than decreased, its relevance to western societies and to societies in transition from previously centralised systems.

The suggestion that Keynes's political thought may be characterised as socialist is normally met with total scepticism. The long-standing majority view in the literature is that whatever kind of political being Keynes was, he was certainly never a socialist or even a quasi-socialist. Usually his position is depicted as either centrist liberal, right-wing liberal or even humane conservative. As the traditional account has it, his goal was to save capitalism from itself to ensure its long-term survival, not to lay the foundations for an alternative type of society.

While the traditional view has always been inadequate, nowadays we have even more reason to criticise it. It primarily took root in the 1950s and 1960s when interpretation was conditioned by two major factors, one internal and one external. The internal factor was the limited knowledge base concerning Keynes's writings and thought. The *Collected Writings of John Maynard Keynes* (1971 – 89) had not appeared, there was next to no interest in his unpublished writings, and there was virtually no awareness of his philosophy. The expanded knowledge base now available reveals the traditional view to be quite deficient.

The external factor was the Cold War. Its ideological currents promoted the line that, in theory, socialism meant Marxism and that, in practice, it meant class politics directed towards the revolutionary defeat of capitalism. Anyone rejecting Marx, permanent loyalty to the working class, and the overthrow of capitalism, could not on this account be associated with socialism. On the simplistic, one-dimensional scale running from Marxists and communists at the extreme left to fascists and reactionaries at the extreme right, it was only natural to view Keynes's thought as falling somewhere in the centre (or right of centre). Today we are not so constrained by the distortions and presuppositions of the Cold War and can assess matters in a more considered light.

The chapter is structured as follows. Some preliminary issues are initially addressed in the first section to clarify certain aspects of the larger argument. The next section presents a long (but incomplete) survey of Keynes's advocacy of liberal socialism, in concept if not always in name, from the mid-1920s to the mid-1940s. This summarises much of the documentary evidence underpinning my argument, but certainly not all of it; further evidence has been presented elsewhere,[2] while other evidence has been entirely omitted for reasons of length. Various features of Keynes's espousal which distinguish it from orthodox socialist advocacy are then outlined. This leads to discussion of the philosophical foundations of his politics and the ways in which these (especially his theory of practical reason) create some of the distinctive features of his approach to socialism. The penultimate section sets forth some of the leading characteristics of his socialism interms of what it is and what it is not. The paper concludes with appeals aimed at rational, constructive debate.

PRELIMINARY ISSUES

Socialism is a family of many variants

This elementary proposition has been lost to popular consciousness as a result of the dominance of Marxism and of ideological simplification generated by the Cold War. As every serious student knows, diverse doctrines of socialism have been propounded over the last two centuries by Babeuf, Blanc, Cabet, Fourier, Lasalle, Marx, Morris, Owen, Saint-Simon and Sismondi, to name but a few. In Britain, Fabian socialism, guild socialism, Christian socialism, Marxist socialism and democratic socialism all took root. Marx (1847) himself acknowledged plurality within the socialist tradition. Though its object was to assert the superiority of 'scientific socialism', the *Communist Manifesto* recognised seven additional forms of socialism, criticism of which did not lead to their expulsion from the socialist family.

It is a central weakness of the traditional account of Keynes's politics that it does not distinguish between different forms of socialism, and thus slips into massive conflation. It presupposes that whenever Keynes criticised or opposed a particular type of socialism he was therefore reacting against socialism in general. Nothing could be more mistaken, and it is one of the objectives of this paper to destroy this myth. The antechamber to deeper insight into Keynes's politics is the recognition that opposition to certain types of socialism does not imply blanket opposition to all types. It is true that Keynes does not always explicitly draw the necessary distinction in his writings, but once one is aware of his advocacy of a particular form, it is a simple matter in reading him to make the distinction for oneself.

A final requirement in debate is consistency of treatment. The criteria applied in judging Keynes's thought should be no different from those applied to all other doctrines generally accepted as socialist.

Thought, life and beliefs

An obvious distinction exists between a writer's intellectual contributions and the conduct of his or her personal life. The two may be interconnected, but the interpretation of their ideas should, as far as possible, be grounded in their intellectual output and not their conduct, especially when that output is rich and copious. In this chapter my prime concern is with Keynes's intellectual system, not his personal behaviour. This means that many aspects of Keynes's life, whether apparently pro-socialist or anti-socialist, are irrelevant to the present argument. The question of the extentto which Keynes's lived life exemplified socialist principles is thus not addressed in this paper. But if it were, the central issue would be conformity of his life with his own understanding of socialism

and not some other. Evidently it would require, as precondition, an adequate account of his version of socialism.

A further obvious distinction exists between intellectual contributions and the author's beliefs in these over time. Once enunciated, ideas have a life of their own separate from the subsequent beliefs of their enunciators. Provided they are internally consistent, a series of ideas advanced at different times can be welded into a system that is independent of any changes in their author's allegiances at a later time. I happen to think that Keynes's political beliefs from the 1920s onwards display temporal continuity in the sense that they belong to the same framework of political philosophy. The unqualified truth of this proposition, however, is not crucial to my central argument. The contention, for example, that a rupture in Keynes's political philosophical beliefs occurred late in his life, say sometime after the *General Theory,* would clearly create important subdivisions within the temporal development of his thought, but it would not affect the interpretation or relevance of the pre-rupture writings.

DOCUMENTATION OF KEYNES'S ESPOUSAL

During his formation as a philosopher and economist up to 1914, Keynes was extremely interested in political questions, both intellectually and practically.[3] His outlook at this stage was that of progressive liberalism, that is, liberalism which recognised weaknesses in *laissez-faire* and accepted a necessary role for the state. Although interested in the discussions of the Fabian socialists, he did not associate himself with any form of socialism. Nevertheless there were, as often occurs with Keynes, some hints of what was to come. His 1909 notes on 'The Relation of Economics to Ethics' referred, for example, to 'Webb's Socialism' and suggested that the ethical justification of wealth meant that the present organisation of society might only be transitional.[4]

It was after World War I that a significant shift in the centre of gravity of Keynes's political philosophy appears to have occurred. The aftermath of the ordeal saw Britain and the world gripped by economic malaise and new political struggles. Keynes's attention now turned explicitly to socialism without, of course, abandoning his liberalism. His objective was a confluence of liberalism and socialism that appropriated what was most worthwhile in each. During the 1920s and 1930s, espousal of what he was later to call 'liberal socialism' became a regular if untrumpeted theme. And while war and postwar activities dominated in the 1940s, his writings of this period do not indicate any fundamental departure from the frameworkdeveloped between the wars. Taken in chronological order, the main supporting evidence is as follows.

The 1920s

In 1923, Keynes's intellectual dissatisfaction with prevailing political ideologies was evident. On the left, communism was 'discredited by events' and 'socialism, *in its old-fashioned interpretation,* no longer interests the world' (KCW/XVII: 450, emphasis added) while, on the right, *laissez-faire* and individualistic capitalism were quite inadequate (KCW/XIX:149). Something different was needed. A hint lay in the above qualification to the term socialism, for it suggests that he may have begun to think, as early as January 1923, that what *might* interest the world was a *new-fashioned* interpretation of socialism.

About sixteen months later, on 24 May 1924, Keynes publicly aligned himself with socialism for the first time. In 'Does Unemployment Need a Drastic Remedy?' he outlined the way forward in terms of a *'true socialism of the future'* (emphasis added). The next stages of 'politico-economic evolution' would lie in 'co-operation between private initiative and the public exchequer', with this type of socialism springing from 'an endless variety of experiments' aimed at discovering 'the respective appropriate spheres of the individual and of the social' and the ways these 'sister instincts' could be fruitfully allied. *Laissez-faire* had had its day, and the 'ultimate cure' for unemployment was to be found, not by crushing organised labour and cutting money wages, but by stimulating investment activity through 'state-encouraged constructive enterprises' and monetary reform (KCW/XIX: 219 – 23).

The article provoked considerable comment to which Keynes penned various responses, the main one being his 'Reply to Critics' of 7 June. While these largely focused on economic issues, general ideas were not ignored. He repeated his abandonment of *laissez-faire,* his call on 'the State', and his belief that 'the next developments of politico-economic evolution will emerge from new experiments directed towards determining the appropriate spheres of individual and of governmental action' (KCW/XIX:228 – 9), all of which were central elements of what he saw as the socialism of the future. Rather than accepting the outflow of British tax revenue to finance infrastructure overseas, the state should redirect these funds to domestic development through a policy of organising, with private sector assistance, large capital works. These would include 'mass-production of working class houses', road building, port construction and electrical power transmission, projects of large cost and moderate expected profit which the private sector would not undertake unaided.[5] The irony for Keynes was that the new Labour government under Ramsay MacDonald was so anxious to avoidaccusations of socialist tendencies that it clung more tenaciously to *laissez-faire* than the Tories (KCW/XIX:223 – 32).

But an even more telling event occurred outside the public gaze, from which it has remained hidden until recently. On 8 June 1924, the day after his 'Reply to Critics' appeared, Keynes drafted an outline for a book with the striking title, *Prolegomena to a New Socialism*. This sketched out his conception of a new kind of socialism adapted to modern conditions. It rejected individualism and

laissez-faire (without rejecting various of their valuable features such as risk-taking), gave the state a wide range of preoccupations, and grounded itself philosophically in the ethical goodness of the whole and a recognition of both benevolence and pleasure in human life.[6] Just as various liberals at the turn of the century had called for a 'New Liberalism', so Keynes now called for a 'New Socialism' which, having cast off the doctrinaire slogans of the past, could meet the challenge of changed historical conditions.[7] Later that year, in November, the fruit of further work on his position emerged—his first lecture on 'The End of *Laissez-Faire*' at Oxford, and a draft outline of a second unwritten book, 'The Examination of Capitalism'—to both of which I return below.

In March 1925, Keynes resumed his socialist theme, prompted by debate on monetary reform, one of the planks of his attack on unemployment.[8] Discussion of Britain's imminent return to the gold standard led him to write to *The Times* to emphasise another key element in his conception of socialism—semi-autonomous public institutions. Such 'semi-independent corporation[s] within the state' represented 'a type of that *socialism of the future* which is in accord with British instincts of government, and which—*perhaps one may hope*—our Commonwealth is evolving within its womb'(KCW/XIX:347 – 8, emphases added). These corporations, of which the Bank of England and the universities were prominent examples, had two primary characteristics—they worked for the public good rather than private profit, and they enjoyed independence from political pressures except in the last resort when parliament could impose its will. Associated with this was the notion that such institutions could evolve out of previously private and individualistic institutions by a definite, but not inevitable, process of 'self-socialisation'.

In 'The End of *Laissez-Faire*' (of 1924 and 1926) Keynes rejected the dominant ideology of liberalism and argued for new forms of social organisation. The 'lethargic monster' of *laissez-faire* ruled by hereditary right, not by doctrinal merit. Its roots lay in the last two centuries—in older political and moral philosophies, in the needs of business, and in the weakness of alternative programmes (specifically protectionism and Marxian socialism). The present task was *demolition*—to 'clear from the ground' the various general principles underpinning *laissez-faire*—followed by *reconstruction* using the distinction between matters pertaining to society and the state ('the Agenda') and matters pertaining to individual activity ('the Non-Agenda').This division could not be determined *a priori* but was historically contingent and case-specific.[9] Economics and politics were assigned companion tasks. Economists were 'to distinguish afresh the *Agenda* of government from the *Non-Agenda*', while the work of politics was 'to devise forms of government within a democracy…capable of accomplishing the *Agenda*'. The economic principle for the state's agenda was to embrace desirable projects which, because they fell outside the individual sphere, were not performed at all. This involved society 'in exercising directive intelligence' through 'appropriate organs of action', over such issues as currency

and credit, the wide collection and publicity of business data, the scale and distribution of national saving, and population size. His principle in the political sphere was that 'in many cases the ideal size for the unit of control and organisation lies somewhere between the individual and the modern State', that is to say, with 'semi-autonomous' bodies or corporations.

Such proposals were directed towards 'possible improvements in the technique of modern capitalism by the agency of collective action'. In relation to socialism, they represented a rejection of State socialism and the adoption of an evolutionary flexible path which worked with the tendencies of the day. Whilst applauding several aspects of State socialism, his complaint was that it was out of touch with current problems and conditions. In relation to capitalism, he saw nothing in the proposals which was 'seriously incompatible with…dependence upon an intense appeal to the money-making and money-loving instincts of individuals'. This, however, was an aspect of capitalism that Keynes despised on philosophical grounds, however much he accepted it as an effective motive force in economics. The essay concluded by posing the key problem in terms of economics and ethics. Given that a wisely managed capitalism is the most efficient system we have for attaining economic goals but that it is in itself extremely objectionable in many ways, our task is 'to work out a social organisation which shall be as efficient as possible without offending our notions of a satisfactory way of life'. The next step in this direction was not political agitation or premature experimentation, but 'thought', 'a new set of convictions' which, based on concordance between intellect and feelings, would provide a platform for reformers to pursue 'right aims by right methods' (KCW/IX:287 – 94).

In 1926 Keynes privately returned to another project first treated in 1924, the draft outline of an unwritten book entitled 'An Examination of Capitalism'. Its tripartite structure of 'Ideal', 'Actual', and 'Possible' revealed much about his understanding of ends and means. Part I focused on the nature of a future ideal society, including distributional justice and the role of money. Part II returned to the present reality of contemporary capitalism, its advantages (mainly efficiency) and disadvantages (mainly ethical but also some inefficiency). Part III, treating of the possible, provided the link between the two by dealing with feasible strategies of transition. Such strategies aimed to preserve the advantages of capitalist efficiency during theprocess of moving to the new society which placed greater emphasis on non-economic goods and social justice. The 1926 version of the outline concluded with a comment on the movement towards socialism: 'Since Socialism will be partly inefficient and will have little or no economic margin, it must begin by emphasising its strongest points and taking guard against its weakest'.[10]

Keynes's contribution in 1927 was 'Clissold' (KCW/IX:315 – 20), a review of a work by H.G.Wells, from the large canvas of which he selected the theme of desirable revolutionary change. Wells's views, together with Keynes's own

sympathetic gloss, may be summarised as follows. In the modern world the motive force behind constructive social transformation consists of new ideas (coming primarily from new liberals) combined with existing feelings (contributed chiefly by socialists). Creative intellectuals (drawn from all quarters, including business people and scientists) and a new philosophical 'creed' born out of the ashes of traditional liberalism and socialism, will provide the means to move society 'far, far to the Left'. From his review, it is evident that such a programme had considerable appeal to Keynes.

In 1928 the Liberal Party's report, *Britain's Industrial Future,* appeared, Keynes being a primary author of Book 2, 'The Organisation of Business'. At first sight this might seem an improbable source of socialist material, but such an inference is hasty. The political outlook of Book 2 springs from two propositions—the rejection of the 'distorted' and 'obsolete' choice between the 'Individualism' of the Conservatives and the 'Socialism' of Labour (namely, 'a comprehensive State Socialism') and, secondly, the adoption of an alternative, middle-way approach which recognises and seeks to appropriate the merits of Individualism and State Socialism while avoiding their disadvantages. 'The task', according to the report, was 'one of guiding existing tendencies into a right direction and getting the best of all worlds, harmonising individual liberty with the general good, and personal initiative with a common plan', a conception entirely in accord with Keynes's liberal socialism. Within this general framework (which incidentally focuses more on micropolicy than macropolicy), the report sets out the following proposals for the conduct of business activity.

That many goods and services are provided by both the private and public sectors in modern societies was accepted and welcomed. The primary objective was not to produce drastic changes in this division, but to enhance the efficiency and performance of each sector and the system as a whole so as to promote the public good. With regard to public enterprises, their methods of management needed major overhaul to make them more 'lively and efficient'. Efficiency was to come, not through nationalisation or direct trading by government, but through management by public boards with statutory powers (along the lines of 'semi-autonomous corporations'). With regard to the private sector, small business escaped intervention, but the conduct of large business, due to its tendencies to separate control fromownership in joint stock companies and to monopolistic practices, required considerable change to the legal framework. The two chief weapons to protect the public interest were *publicity* and *supervision*. Compulsory publicity was required for complete company accounts, directors' shareholdings and Board of Trade investigations into company abuses. Greater supervision was to be obtained by establishing a Supervisory Council, representing shareholders and possibly employees, to oversee appointments to, and the performance of, the board of directors of the company, and by giving the Board of Trade open access to the decisions of industry associations to prevent

cartelisation. In extreme cases, the Board of Trade could recommend price controls, profit-sharing and worker representation.

At the macro level, the report sought more efficient mechanisms to direct saving into productive investment. For private investment, abuses of the Stock Exchange were to be removed and efforts made to ensure the right distribution of saving between domestic and foreign investment. For public investment, Keynes renewed his advocacy of a National Investment Board which, as part of Treasury, would draw up a budget of capital expenditure on public works financed out of funds pooled under its control. This Board was viewed as a powerful means of employment and wealth creation involving minimal interference with the private sector. Keynes's section then rounded off with additional initiatives to improve overall efficiency—an Economic General Staff offering systematic analysis and skilled advice to Cabinet, a Committee of Economic Policy chaired by the Prime Minister, the widening and deepening of official statistical information ('The nationalising of knowledge is the one case for nationalisation which is overwhelmingly right'), and various measures to enhance business efficiency as a whole such as increased cooperation, rationalisation, management education and marketing.

The 1930s

During the second Labour government under MacDonald (1929 – 31) Keynes became an important advisor to the Prime Minister and government.[11] In this milieu, as the leading critic of the orthodox views of Treasury and the Bank of England, he advocated policies promoting the government's socialist objectives and furnished immediate plans and longer-term strategies. Chief among these were expansionary employment policy, public works and protection. A revealing incident in the Economic Advisory Council occurred when Keynes sought an enquiry into state alleviation of unemployment but was blocked by the business members and Snowden, the strictly orthodox Chancellor of the Exchequer. According to Dalton (1953: 261) Keynes then described himself as 'the only Socialist present', a remark which wounded MacDonald's feelings.[12]

In December 1930, the spirit of Oswald Mosley's manifesto (circulatedafter his failure to convert the Labour Party to economic expansion) attracted Keynes. It proposed a 'scheme of national economic planning…and wide executive powers' (KCW/XX:473 – 4). As Keynes saw it, policy choice ultimately came down to two basic alternatives: reliance on natural forces and the invisible hand (by either doing nothing or promoting them with money wage cuts), or the adoption of 'a scheme of collective planning', instances of which had been promoted by the Liberal's Yellow Book in the last election, Lloyd George's agricultural programme and the Mosley manifesto. The choice was directly linked to socialism.

> For the choice will be ever more openly set between forcing a reduction in wages and a scheme of national planning. But…looking further ahead *I do not see what practical socialism can mean for our generation in England, unless it makes much of the manifesto its own—this peculiar British socialism, bred out of liberal humanitarianism, big-business psychology, and the tradition of public service.*
>
> (KCW XX475, emphasis added)

He understood why supporters of *laissez-faire* would be antagonistic, but 'how anyone professing and calling himself a socialist can keep away from the manifesto is a more obscure matter'. Speaking in terms of policy positions which cut across normal party lines, Keynes identified three main political 'parties' or groupings—the Socialists, Liberals and Conservatives. It was with the 'Socialist' grouping that his general sympathies lay.[13]

On 13 December 1931, Keynes delivered a lecture to the Labour Forum of a left-wing organisation, the Society for Socialist Inquiry and Propaganda. His chosen topic was 'A Survey of the Present Position of Socialism', the occasion prompting it being Labour's crushing defeat in the general election of the previous October. This lecture is critical in understanding Keynes's form of socialism and the strategy he urged for its attainment. It also clarifies some of the central differences between his form and the hegemonic socialisms of his time. A revised version of the speech was subsequently published in *The Political Quarterly* in 1932 under the new title 'The Dilemma of Modern Socialism'. Although the two documents are quite close, the unpublished speech contains interesting additional material and personal comment absent from the more academic publication. The following account draws on both documents.[14]

Keynes's analysis is written from the standpoint of his own type of socialism, and from that viewpoint it is a brilliant and incisive commentary. It is perhaps the most succinct statement of his overall position. The 'dilemma' addressed was that of *choosing the right programme or strategy for the achievement of socialism*.

Three general motifs were identified, present in various proportions in all socialists and pertinent to various phases in the achievement of socialism:

1 the 'political', referring to the attainment of power,
2 the 'practical', referring to wealth creation and the solution of the economic problem, and
3 the 'ideal', referring to the future society where higher non-economic goods take precedence.

Underpinning his account is a crucial distinction between 'economically sound' and 'economically unsound' methods of increasing wealth and solving the economic problem. Not surprisingly, Keynes's strategy gave priority to the

second motif which involved doing what was 'economically sound' to increase wealth. His objection to giving priority, at the present conjuncture, to the other motifs was that they involved doing what was (or might be) 'economically unsound'—the first through political action which might disrupt or interfere with production, the third through sacrificing further growth in wealth for higher values. However, while stressing the practical motif as important in all phases, he acknowledged that the political motif would precede it in time, understood as a peaceful and democratic assumption of power.

For Marxists, the three motifs were respectively represented as 'The Revolution, the Five Year Plan, the Ideal'. His differences from Marxian socialists were that they laid primary stress in the first phase on the political motif, interpreted as revolutionary change which disrupted the economy and rendered the present system unworkable, and relegated the practical motif to a subsequent phase in which a new economic system would be established. As Keynes saw it, the reduction in resources and impoverishment caused by wrecking the existing system was likely to make construction of a new system very much harder or even impossible, thereby generating an almost certainly worse outcome than present society.

In his speech he outlined his strategy for socialism, and his differences from other socialists, as follows:

> For my part I would define the Socialist Programme thus:
>
> To obtain political power with a view to doing what is economically sound in order that the community may become rich enough to *afford what* is economically unsound.
>
> My goal is the ideal,…to put economic considerations into a back seat, but my method *at this moment of economic and social evolution* would be to advance to my goal by concentration on doing what is economically sound.
>
> But there are others—and it is with these that such as I could never co-operate—who would like to concentrate on what is economically unsound, because they believe—falsely in my judgment—that this is the best way to gain political power (which, we allagree, is the first step) and that to render the existing system unworkable is the only means of reaching a new system…[15]

Two 'good and sufficient reasons' were given for concentrating on the practical motif. One was that economically sound reforms pointed in the same direction as the ideal and not, as in the past, away from it—'the central control of investment and the distribution of income…will also tend to produce a better kind of society on ideal grounds'. The second was that economically sound policies might permanently solve 'the problem of poverty'. Radical changes in techniques of production ('brilliantly' described by Fred Henderson)[16] made possible vast

increases in consumption goods and a diversion of employment to services and capital goods. His emphasis on the practical problem also helped explain his differences with the Labour Party. He saw it as both largely anti-intellectual with no interest in new theories of what was economically sound, and old-fashioned in clinging to outmoded *laissez-faire* policies. It was therefore an obstacle to Keynes's programme until it became intellectually emancipated in relation to what was economically sound, and also in an electoral impasse until it allowed the practical motif to prevail over the revolutionary form of the political motif.

His speech finished with a return to his 'own personal position':

> We stand at a point in economic evolution when it seems to me vital to concentrate on what is economically sound….
>
> I conclude by repeating what should be, in my judgment, *the deliberate and self-conscious aim of the British socialist today*—To gain political power in order to do what is economically sound and thereby make us rich enough to be able to afford to do things which are economically unsound….
>
> Can the Labour Party which now lies stunned and prostrate, learn to employ its political craft and its deep-founded political strength so as to harvest the power of idealistic unselfishness and public spirit…to the task of creating a sound economic engine which will end poverty? (emphasis added)

The final proposition of the article was that 'temporary concentration on the practical is the best contribution which we of today can make towards the attainment of the ideal' (KCW/XXI:38). This urging, doubtless reinforced by his experiences and disappointments as advisor to MacDonald's government, is also grounded on his philosophical position as explained in a subsequent section.

Keynes's identification with a particular conception of, and strategy for, socialism in his speech and article could hardly be clearer. The programme he defined as *the right socialist programme for current conditions,* and which he urged the Labour Party to support, was *his own programme*.

While socialism naturally does not play a major role in the *General Theory (GT),* it is nevertheless implicit, primarily towards the end where political-economic issues are more prominent. In a 1932 draft the final chapter was strikingly entitled 'Socialism' (KCW/XXIX:50).[17] In the 1936 work, socialism left its mark in a number of areas of which four may be noted. The first was the placing of certain 'central controls' in the hands of the state so that the economy could achieve more optimal results. Elsewhere Keynes had called this process 'planning'. The second, a major component of the first deserving special mention because of its notable terminology, was the advocacy of 'a somewhat comprehensive socialisation of investment'. This did not imply state ownership of the means of the production, but state management of the total volume of investment along the lines of his proposals in the 1920s for a National Investment

Board, together with various forms of cooperation between the public and private sectors (KCW/VII:378). Thirdly, the 'arbitrary and inequitable distribution of wealth and incomes' (one of the two 'outstanding faults' of capitalist society) was to be attacked partly by redistributive taxation and partly by lowering the rate of interest to contribute to the 'euthanasia of the rentier' or functionless investor (KCW/VII:372 – 7). Finally, there is Keynes's assessment of Gesell's *The Natural Economic Order*. In opposing *laissez-faire,* Gesell aimed at 'the establishment of an anti-Marxian socialism' founded on a rejection of 'classical' economics,[18] and on an unfettering of competition instead of its abolition. The parallels with Keynes's own thinking, while not complete, are notable. The future 'will learn more', he opined, 'from the *spirit* of Gesell than from that of Marx' (KCW/VII:355, emphasis added).

Post-*General Theory* controversy and his heart attack minimised politically oriented comments during 1937. In 1938, however, he used his chairman's speech to the National Mutual Insurance Company to promote his views on the compatibility of planning and liberty.

> To favour what is known as planning and management does *not* mean a falling away from the moral principles of liberty which could *formerly* be embodied in a simpler system. On the contrary we have learnt that freedom of economic life is more bound up than we previously knew with the deeper freedoms—freedom of person, of thought, and of faith.
>
> (KCW/XII:238 – 9, emphases added)

But it was not until January 1939 that Keynes returned to explicit public advocacy of his form of socialism. During discussion with Kingsley Martin, the left-wing editor of the *New Statesman and Nation,* on the relations between democracy and economic efficiency, he viewed the situation as follows. 'In contemporary conditions we need…so much more central planning than we have at present'. Private capitalism was in 'decline as a meansof solving the economic problem', and 'the only practicable recipe' was a 'particular amalgam of private capitalism and state socialism' (KCW/XXI: 492). This middle-way programme posed no practical threat to personal and political liberty so long as the connection between these liberties and the rights of private property (stressed by *both* liberalism and the French Revolution) were recognised. Most of society, including the bulk of politicians, were actually liberal in their convictions, not in the party sense, but in 'the most fundamental and genuine meaning of the word'. The Labour Party, in particular, was heir to 'eternal liberalism' and not the home of 'demi-semi-Fabian Marxism' (KCW/XXI:494 – 6). He concluded his politico-economic analysis by giving his political views, for the first time, the distinctive name of 'liberal socialism'.

> The question is whether we are prepared to move out of the nineteenth century *laissez-faire* state into an era of liberal socialism, by which I mean a system where we can act as an organised community for common purposes and to promote social and economic justice, whilst respecting and protecting the individual—his freedom of choice, his faith, his mind and its expression, his enterprise and his property.
>
> (KCW/XXI:500)

It deserves repeating that it was in 1939, well into the later years of his life, that Keynes placed 'liberal socialism' on the banner flying over his political position.

The 1940s

During the Second World War, Keynes's schemes were underpinned by the same political-economic framework and the same concern for economically sound policies. In 1940, his proposal for deferred pay in *How to Pay for the War* was related to broader considerations of regulation and liberty.

> I am seizing an opportunity…to introduce a principle of policy which may come to be thought of as marking the line of division between the totalitarian and the free economy…. Just as in war the regulation of aggregate spending is the only way to avoid the destruction of choice and initiative,…so in peace it is only the application of this principle which will provide the environment in which the choice and initiative of the individual can be safely left free. This is the one kind of compulsion of which the effect is to enlarge liberty.
>
> (KCW/XXII:123)

In 1942, he described his plan for postwar commodity policy as aiming at 'a middle course between unfettered competition under *laissez-faire* conditions and planned controls which try to freeze commerce into a fixed mould', the latter referring to 'state trading on Russian lines' (KCW/XXVII: 111).

In 1943 he concluded his remarks on full employment policy which depended *(inter alia)* on a stable long-term public investment programme, by endorsing Henderson's comments on socialism which sought a *modus vivendi* in which the state managed the economic system as a whole without owning, controlling or interfering in particular businesses (KCW/XXVII:320 – 6).

In 1944, on postwar employment policy, he supported the view of those economists

> who believe that it will be the role of this country to develop a middle way of economic life which will preserve the liberty, the initiative and (what we

> are so rich in) the idiosyncrasy of the individual in a framework serving the public good and seeking equality of contentment amongst all.
>
> (KCW/XXVII:369)

And finally, in his Lords speech of December 1945 shortly before his death, he observed in relation to the Anglo-American loan negotiations that

> the outstanding characteristics of the plans is that they represent the first elaborate and comprehensive attempt to combine the advantages of freedom of commerce with safeguards against the disastrous consequences of a *laissez-faire* which pays no direct regard to the preservation of equilibrium and merely relies on the eventual working out of blind forces.
>
> (KCW/XXIV:621)

The above instances deal with Keynes's explicit espousals of his type of socialism, or aspects of it, from the 1920s to the 1940s. But many of his other writings which, while not specifically mentioning socialism or even politics, nevertheless represent elaborations of the total picture. These deal, for example, with the possibilities offered by future societies, the production and consumption of non-economic goods, and the importance of morality in the conduct of life.

SOME FEATURES OF KEYNES'S ESPOUSAL

Let us now take stock by noting some of the main features of Keynes's espousal of socialism. Firstly, the espousal is *steady,* durable and unwavering;he advanced his views in one form or another in virtually every year from 1924 to his death. Secondly, it is *clear,* explicit and unambiguous; he used the term socialism to characterise his own views, he addressed the problems of socialism as he saw them, and he provided strategies for the achievement of his version of socialism. Thirdly, it is *public* and not merely private, appearing in a variety of journal articles, *The Times* newspaper, and a BBC radio broadcast.

Finally, it is *quiet* in tone. There is none of the stridency, drum-beating, calls to battle or frequent declarations of allegiance that accompany certain other socialisms. Quietness is a consequence of his political philosophy and position of influence. As noted below, his socialism is characterised by cooperation rather than conflict, persuasion rather than force, and reform rather than revolution. His primary arena was that of the intellect rather than daily politics, and to effect change he worked to influence the ideas and opinions of those in positions of power. A constant pushing of a socialist line was likely to be misunderstood, resulting in a loss of persuasiveness within the establishment. Thus, quite apart from the view that political philosophy is not to be derived primarily from personal conduct, the fact that Keynes's behaviour did not conform to that of traditional socialists is perfectly explicable by his own political philosophy.

By way of further clarification, suppose a critic sympathetic to socialist principles perceives theoretical or practical weaknesses in the dominant form(s) of socialism and proposes an alternative. The intellectual response, as distinct from the emotional one of labelling the critic non- or anti-socialist (because of a conflation of particular and general forms) is to engage in open-minded discussion and further analysis. Suppose one of the critic's claims is that insufficient attention is being paid to the *preconditions* of successful socialism. The currently dominant form, it might be argued, is prone to collapse because its economic organisation is unlikely to generate an adequate standard of life, because its political organisation is excessively bureaucratic and oppressive of individual freedoms, or because it undervalues the importance of standards of morality in public life. It would be a sign of a serious socialist to advance the critique and to wrestle with the construction of a new kind of socialism, one which gave priority to laying the (economic, political or moral) foundations of successful transformation rather than striving merely for the attainment of political power and the overthrow of the existing system, the likely outcome of which could be much worse than present society.

Such, I contend, was Keynes's project. Dissatisfied with existing versions of British socialism and central elements in the Soviet system, he set out constructively to formulate a new socialism which aimed at durable foundations, steady forward movement as far as current circumstances permit, and programmes with real prospects of practical success.[19]

It is undeniable that Keynes did not think or act like many other socialistsof his period. But this does not constitute a reason for denying his advocacy of socialism, because his form of socialism is philosophically different from other forms. To gain deeper insight into his differences from traditional socialisms, we now turn to the philosophical foundations of his politics, particularly his ethics and philosophy of practical reason.

PHILOSOPHICAL FOUNDATIONS

The greatest and most enduring philosophical influence on Keynes was G. E.Moore's *Principia Ethica*. Although Moore himself had practically no involvement in public affairs, he advanced a philosophical framework which embraced politics at a general level and within which Keynes's particular politics moved. Keynes made a number of significant modifications to Moore's ideas, but these did not overturn the general system. It was a means-ends framework in which the ultimate end of rational human action was the generation of as much intrinsic goodness as possible, and in which the means embraced all the sciences, politics and economics included. A number of important political corollaries followed.

Firstly, ethics is the master, and politics and economics the servants. Politics and economics can only provide truths about means, never about final ends upon

which ethics alone is able to pronounce. While political or economic systems or sets of principles may be variously good as means, they do not possess intrinsic goodness or constitute ends in themselves.[20]

Secondly, political action is based on the theory of practical reason. It is therefore to be directed towards creating wholes of as much ethical value as possible. In this domain, Keynes took over Moore's consequentialism but replaced Moore's treatment of uncertainty with his logical theory of probability, thus arriving at the conclusion that the best humans could do under conditions of limited knowledge and' foresight was to aim at wholes which increased the amount of *probable* goodness in the world, *ceteris paribus*. This reconstruction had the important effect of liberating practical ethics and hence politics from the predominantly conservative state in which Moore had placed it. Politics no longer needed to be linked so closely to the *status quo,* but could envisage schemes of radical reform and change.

Thirdly, the determination of right courses of action by means of consequentialist ethics applied to organic wholes implies the dependence of rational action on *circumstances*. On this view, there are no principles, rules or duties of a purely political kind that ought universally to be followed in practice irrespective of time and place. It is thus inconsistent with this philosophy of practice to say that socialism, for example, always and everywhere requires *x,* where *x* might be 'public ownership of the means of production', 'nationalisation', 'siding with the proletariat in the class struggle', 'economic equality', and so on. Whether such objectives areappropriate at a given conjuncture is not to be decided by appeals to universal propositions, but by consequentialist assessment which takes account of relevant known circumstances. The nature of Keynes's philosophy of practice and its failure to sanction universal political maxims goes a long way towards explaining key differences between his thought and action and those of traditional socialists.

Fourthly, considered as means, politics and economics cannot establish ethical goodness directly, and can only create the *preconditions* of goodness. These preconditions facilitate increases in goodness by removing barriers to the pursuit of intrinsic goodness by individuals, and by actively encouraging such pursuit. The preconditions include the protection of life and property, individual freedom, adequate material prosperity for as many individuals as possible, economic efficiency, peace, a right population trend, concern for the natural and cultural environments, cooperation and social contentment (including the reduction of high economic inequalities). It was their critical importance as preconditions that led to Keynes's stress on practical economic matters as central to the achievement of sustainable socialism—economically sound policies, efficient production, and high employment in a context acceptable to notions of social justice. A constant thread in Keynes's thought is the proposition that durable and acceptable socialisms can only be constructed on sound economic

foundations, that is, within sufficiently efficient economies with adequate resources and wealth.

Keynes's ultimate end or 'ideal' was a society ruled by ethical rationality; that is, a Utopia characterised by increasing quantities of intrinsic goodness. Such a Utopia had no endpoint of perfection, for goodness was something capable of indefinite increase. The chief intrinsic goods were the mental states of love and personal affection in their many forms, and of the aesthetic enjoyment of beauty in all its forms. These could only greatly flourish and expand once the relevant preconditions were in place.

In the transition to the ideal, Keynes saw two main obstacles—one 'economic', the other 'psychological or moral'. With his typical optimism, he thought the economic problem could be relatively easily solved, in the sense of providing everyone with a minimum level of economic well-being. The moral-psychological problem was far more complex and difficult, however. It required defeat of the two oppressions of 'money-love' and 'purposiveness' characteristic of capitalist society. The acquisitive obsession with money for its own sake and the postponement of enjoyment to a never-arriving future had to be rolled back so as to allow individuals to appreciate and enjoy the higher non-economic goods in the present. Eliminating the dominance of money-love and purposiveness required major changes in psychology and values. In the short term, human nature could be managed but, in the long, transmutation was required.

The fundamental questions of politics thus centred on processes of evolutionary change in which capitalist society, with all its merits and faults,gradually acquired the relevant preconditions, and was progressively transformed into a liberal socialist, ethically rational society.

Burke's influence relative to Moore's

Some commentators, such as Robert Skidelsky (1983, 1992, 1996), view the conservative philosopher Edmund Burke as far more fundamental to Keynes's politics than Moore. Using an opposition between Keynes's private and public lives, Skidelsky argues that while Moore influenced Keynes's private life, it was Burke (and a spirit of public service as exemplified by his mother) that formed his politics. One reason for this, according to Skidelsky, is the presence of a 'mixed goods dilemma' in Moore's ethics which blocks would-be Moorean reformers such as Keynes from pursuing large or radical schemes for social change and confines them to small and modest proposals. The result is a quite politically conservative account of Keynes's politics completely at odds with the account presented here. I have shown elsewhere that the 'mixed goods dilemma' is entirely fictional, existing in the thought of neither Moore nor Keynes (O'Donnell 1998). Here I should like to reemphasise the main reasons why reliance on Burke rather than Moore as the key to Keynes's politics is mistaken.

My view is that Burke did exercise a degree of influence on Keynes, but it was subsidiary and not fundamental. Burke is secondary to Moore for the following reasons:

1 Keynes came under the spell of Moore's *Principia Ethica* in October 1903, well before he wrote The Political Doctrines of Edmund Burke', which was completed in November 1904.
2 The criteria Keynes used in evaluating Burke's politics were derived from Moore's ethics. If one examines the parts of Burke that Keynes admired and the parts he rejected, they fall quite neatly into the pattern one would expect from a Moorean philosopher. And many of the rejected parts are fundamental to Burke's conservatism.
3 Keynes's theory of practical reason, upon which his political philosophy rests, is totally independent of Burke's thought. Not a single element from Burke is needed or used to. establish this theory. The elements derive primarily from Moore and Keynes, with contributions from Russell and probability theorists such as D'Alembert. The theory, set out in the fellowship dissertations and the *Treatise on Probability,* could have been constructed if Keynes had never read Burke.
4 Too many inconsistencies are created with Keynes's subsequent writings, inconsistencies which stem from the attempt to ground the thinking of someone constantly aiming for reform and social progress in a variety of spheres on the thought of a political conservative.

SOME MAIN CHARACTERISTICS OF KEYNES'S SOCIALISM

To assist in clarifying some of the features distinguishing it from other socialisms, I shall first provide an indication of what Keynes's socialism is *not,* before passing to what it *is*. No attempt is made to be exhaustive, however, the scope of the subject exceeding the space available by a large margin. Keynes's socialism is *not* founded on the following ideas or principles:

Marxism. Keynes saw Marxist doctrine as fundamentally 'illogical' (KCW/IX: 285) and attributed to it the same intellectual roots as classical *laissez-faire* economics. Keynes's view of Marxism is easily criticised, but the fact that his understanding of Marx is inadequate does not constitute grounds for rejecting him as an advocate of a different form of socialism. The same fault could be laid at the door of various other socialists.

Class. Class was never a fundamental category in Keynes's conceptual framework, though it could, and did, play a role in his analysis from time to time. He accepted that class existed and possessed some importance, but it was a secondary influence dependent on circumstances rather than a primary and permanent influence. It was never determining in the final analysis, this role, by

and large, being attributed to ideas. In terms of party politics, his aspiration was for a non-class party which transcended traditional class divisions.

Revolutionary upheaval. On Keynes's theory of practical reason, it is impossible abstractly to sanction revolutionary upheaval in which the large evils of the present and near future are virtual certainties and the great goods promised for the distant future are very doubtful. Probabilistic consequentialism indicates that the balance of advantage will typically lie with evolutionary rather than revolutionary courses of action.[21] However, Keynes's vision combined significant change in the short- and medium terms with major social transformations in the long. In this regard, his thought may be summed up in the slogan: 'short-term reform and long-term revolution'. But whatever the pace of change, whether rapid or measured, it was to be peaceful, democratically acceptable, and under social control as much as possible.

Distributive criteria. Keynes's socialism is not defined in terms of abstract rules or criteria concerning distributional equality or 'social justice'. In his theory of practical reason, these issues are important elements in the whole situation but are never the only or hegemonic elements. Agreement with political thinking grounded solely on distributive issues was therefore impossible.

State socialism. Although strongly supportive of his own conceptions of state planning and central supervision, Keynes always opposed central planning and central controls over the economy as a whole.

Public ownership of the means of production. No matter how public ownership is conceived, it was never one of Keynes's guiding principles. What mattered far more was performance (in economic and social terms) rather than ownership. The state would typically own or be involved with a range of productive assets, but calls for nationalisation as matters of principle were irrelevant. Each case was to be considered on its merits, but in general he accepted the considerable size of the public sector of his day and saw no general need to increase the area of the economy under public ownership.

Let me now pass to the positive delineation of Keynes's type of socialism, commencing with general remarks before turning to more specific points.

It was 'socialist' because society, acting through the state, exercised a degree of conscious control over its material and spiritual destiny. The state's role was that of the guardian and promoter of an economically and ethically rational society, a role which placed a wide range of issues on its agenda. When economic and ethical imperatives diverged, it was a matter of devising priorities within a framework embracing both the short and long term. As various of his writings demonstrate, this generally meant for Keynes an emphasis on the economic and material foundations in the short term, preparatory to the flowering of the ethical in the long.

It was 'liberal' because it insisted on protecting and nurturing individualism. Not individualism in the uncontrolled or *laissez-faire* sense which Keynes constantly criticised, but an individualism which sought as broad a field for

personal freedoms as possible in the pursuit of increasing intrinsic goodness. Freedom of thought and expression were often first on the list, but freedom of enterprise and private property were not far behind. In fact, provided both were rightly conceived, Keynes viewed the modern era as one in which the broad currents of individual freedom and enterprise on the one hand, and state planning and control on the other, were mutually supportive rather than antagonistic.

While rejecting *laissez-faire* capitalism and state socialism as global frameworks, Keynes sought a fusion or 'amalgam' in which the best features of each could be retained and the worst discarded. What private capitalism had to offer was greater productive efficiency, achieved primarily through free enterprise, incentives for risk taking and the deployment of decentralised units. Its main faults were that it was blind or anarchic, did not guarantee economic efficiency and, through its ethical indifference, was morally objectionable in a variety of ways. What state socialism offered, on the other hand, were the principles of society guiding its own destiny, of central supervision of the system as a whole, of experimentation, and of thepossibility of greater ethical rationality. Its excesses were relative economic inefficiency, rigid centralisation and the suppression of various personal freedoms.

In the development of society along liberal socialist lines, an *experimental attitude* toward change is vital, in which new schemes can be attempted, programmes modified, and mistakes corrected. Courses of action decided by *ex ante* consequentialism can only be recommended on the basis of probable or presumptive advantage because of our ignorance of the future. They may turn out *ex post* to be successful, inconclusive or a failure. The costs of any foreseeable mistakes, moreover, should not be prohibitive or devastating when things go wrong. The nature of change thus needs always to be measured and self-correcting.

For Keynes, rational social change was generally *evolutionary* rather than revolutionary in character, and to that extent it had a sensibly conservative streak. Where possible, change should seek to preserve and respect the things of value bequeathed by the past, and not risk or destroy everything in the revolutionary overthrow of existing society.

Keynes's writings nowhere contain an abstract defence of democracy, or indeed of any political system, as eternally correct. Nevertheless, in tenor and thrust, his thought is highly supportive of *liberal democracy*. It flows naturally from his philosophical outlook, his individualism, his emphasis on freedom of expression, and from the need for workable change to be acceptable to the population. He was also vehemently opposed to totalitarianism, chiefly because of its persecution of individuals and its suppression of personal freedoms.

As regards *social justice* and the distribution of income and wealth, Keynes's position was very similar to that of other practical socialists. These issues were always important elements in politics and economic policy but were not the only elements. To illustrate their importance in Keynes's thought, three examples may

be cited: the promotion of 'social and economic justice' entered into the socialist side of his 1939 definition of liberal socialism (KCW/XX:500); the 'arbitrary and inequitable distribution of wealth and incomes' was depicted in the *GT* as one of the two 'outstanding faults' of contemporary capitalism (KCW/VII:372); and the 1931 Budget and Economy Bill were strongly attacked as 'replete with… injustice' (especially towards schoolteachers and other government employees) and as failing to pursue national policy objectives with methods that respected 'the principles of social justice' (KCW/IX:145 – 6). He was even more passionate about the 1931 Budget in a speech to MPs: 'I should like to say a word first of all on social justice. Can scarcely trust myself to speak. The attacks on the schoolteachers seem to me to be a most foul iniquity' (KCW/XX:608). But such issues were not the sole consideration, and had always to be conjoined with other factors. As a result, the argument always came down to one about *degrees of inequality under thecircumstances*. Some inequality was accepted as inevitable for a capitalist economy laying foundations for the future, although he thought that the degree of inequality in the 1930s was excessive and ought to be reduced. In the longer term, reductions in economic inequality could be addressed more squarely and, although some inequality might continue, it would cease to be an important social concern as people focused their attention on higher non-economic goods. But while the general thrust of his thought was clearly to promote social justice and reduce inequality over time, his politics and economic policies always referred to other matters as well. The presence of these other matters, by partly obscuring his concern for social justice, seems to be one reason why many commentators have downgraded the significance of social justice in his thought.

That the state should control the level of investment in the interests of macroeconomic efficiency was a key proposition of the *GT,* although public investment had been a key factor in Keynes's thought long before 1936. Taken as a whole, his scheme for public sector investment comprised both central and decentralised agencies. At the centre, the level was controlled by the state through a National Investment Board (NIB), the main task of which was to have pre-prepared schemes of public investment ready for implementation in case of shortfalls in investment decided by private and public enterprises. The work of the NIB was to be supported by extensive data collection and processing organisations. The decentralised satellites of the public sector were the 'semi-autonomous corporations' responsible for managing various public activities. These would decide their own investment plans but could presumably be used, when appropriate, as vehicles for NIB projects.

Few, if any, commentators have advanced a definition of the 'socialisation of investment' advocated in the *GT* (KCW/VII:378). It clearly does not mean state ownership of the instruments of production or nationalisation, but beyond this Keynes is not very specific. The definition I propose is that it means bringing the total volume, and partly the direction, of investment under social control, where

social control means state supervision. It is necessary to add the phrase about direction because investment takes specific forms and one cannot normally control volume without affecting direction. This definition is consistent with Keynes's writings on investment before, during and after the *GT,* including his remark that he expected 'to see the State…taking an ever greater responsibility for directly *organising* investment' (KCW/VII:164, emphasis added).

FIVE APPEALS

For purposes of constructive, rational discussion, I should like to make five appeals. The first is that Keynes's espousal of socialism be taken seriously indebate over his political philosophy. Keynes was both highly intelligent and sincere in his beliefs. However much one might disagree with them, one does not doubt (without independent evidence) that he meant what he said in his writings. That virtually every year between the mid-1920s and mid-1940s, in a variety of public fora, he distinctly aligned himself with his version of socialism is obviously not something to be ignored. Yet the case for considering Keynes's thought as socialist, although advanced some years ago, has been greeted with total silence by virtually all commentators.[22]

The second is that issues should be decided on the basis of the evidence available and not with reference to preconceptions or appeals to secondary authorities. The primary (but not exclusive) source of evidence is constituted by Keynes's own writings, both published and unpublished. I interpret these as consistently supporting the view that Keynes advocated a particular form of socialism over approximately two decades. But whether right or wrong, this and all competing propositions should be judged by the tribunal of Keynes's writings. If past experience be any guide, this will still allow a range of differing interpretations.

My third appeal is for consistency of treatment of members of the socialist family. If one kind of socialism is admitted to the fold, then the same criteria must apply to others. Whatever, therefore, permits membership to Christian socialism, for example, should also apply to liberal socialism, Marxist socialism or any other variant. In particular, mere difference from hegemonic forms of socialism does not constitute acceptable grounds for exclusion. If the hegemonic form is rejected and a new form advanced, what matters in this context is not the challenge to orthodoxy but the nature of the putative new form. The fact that Keynes rejected nationalisation (a central feature of one variety of Labour socialism) or class struggle (a key element in Marxism) for instance, is irrelevant unless it is the case these notions are essential to all forms of socialism.

My fourth appeal, consequent to the third, is for further research into the general concept of socialism with a view to the development of satisfactory working definitions. This applies to antagonists as well as protagonists in the debate. For the conversation to rise above ' 'Tis, 'tisn't', we need to know the

senses in which the term is being used. Only then can we determine whether a proposed definition embraces all or only some recognised socialisms, and whether a putative form of socialism should be regarded as a 'full' or only a 'tendential' socialism.

My final appeal is for political tolerance. Socialism has evolved and will doubtless continue to do so. Whichever form is currently dominant is not the last word on what socialism is or must be. (In any case we could not possibly judge whether it was the last word because we lack a clear general conception of socialism.) In recent decades, the political left has suffered crises of confidence and direction, to which the response of thoughtful intellectuals on the left, including some Marxists, has been to explore theoreticalframeworks capable of combining socialism and liberalism. Keynes journeyed the same road from the 1920s to the 1940s as these modern thinkers are travelling in the 1980s and 1990s. Each has his own style of travel due to differences in philosophical vehicle and theoretical luggage, but the general directions and purposes are similar.

ACKNOWLEDGEMENTS

I am indebted to Peter Groenewegen, Craig Freedman and Peter Kriesler for helpful comments, and to King's College, Cambridge for permission to quote from Keynes's papers.

NOTES

1 It gives me great pleasure to contribute this paper because it combines two central influences on Geoff Harcourt's economic and political beliefs—Keynes and socialism (on which see, for example, Harcourt 1986, 1996). An earlier version was presented to the 1991 Australian Conference of Economists at the University of New South Wales where it received kindly mention in Geoff's after-dinner speech.

2 See O'Donnell (1989: chs 13, 14; 1991, 1992). The present paper is a continuation of these earlier contributions.

3 He debated political issues at the Cambridge Union, he electioneered in Liberal campaigns in Birmingham and Cambridge in 1910, and in 1911 toured Ireland with the Eighty Club and was converted to Home Rule.

4 Unpublished notes for the Political Economy Club, Keynes Papers, UA/6/15, King's College, Cambridge.

5 The other benefit of the policy was that by reducing capital outflow, it relieved downward pressure on the exchange rate and upward, policy-induced, pressure on the interest rate (KCW/XIX:233 – 7).

6 Keynes's outline is reproduced and discussed in O'Donnell (1992:781 – 5, 806 – 7).

7 For a discussion of 'New Liberalism' and its relations to Keynes's political thought, see O'Donnell (1989:316 – 21).
8 His chief proposals here, for price stability and adequate credit, were continuous with his *Tract on Monetary Reform* of December 1923.
9 This approach was consistent with Keynes's philosophy of practice (some characteristics of which are mentioned in later sections); for more detailed discussion, see O'Donnell (1989: ch. 6).
10 For the outlines of the proposed work and further discussion, see O'Donnell (1992: 785 – 93, 807 – 12).
11 He was appointed in November 1929 to the Macmillan Committee enquiry into finance and industry, in January 1930 to the Economic Advisory Council (EAC) chaired by MacDonald which pioneered regular economic advice to government, and in July 1930 to the chair of a smaller Committee of Economists when consensus in the EAC proved elusive.
12 For further discussion of Keynes's relationships to Labour governments and Labour socialism, see Durbin (1985).
13 Keynes's support for this practical, British socialism is not contradicted by his remark that he would 'gladly entrust the fortunes of this country sometimes to one of [these parties] and sometimes to another, according to the appropriate pace and immediate direction of progress' (KCW/XX:476). As his qualification makes clear, choice of government depends, not on eternal loyalty to party, but on the existence and nature of progress, and hence on the ideas, policies, leadership and other circumstances of the time.
14 The *Collected Writings* treatment of these writings cannot be described as satisfactory. Only Keynes's published article of 1932 is reprinted (KCW/XXI:33 – 8) and not the 1931 speech which in certain respects is more illuminating (Keynes Papers PS/5). It also tucks away a quintessentially political article between two economic pieces in a long chapter on 'The currency question' in an *Activities* volume devoted primarily to economic matters *(Activities 1931 – 9: World Crises and Policies in Britain and America),* rather than presenting it in a volume or chapter dealing with politics. In its accompanying headnote, it provides little contextual information and does not give the full name of the society Keynes addressed.
15 Emphasis in original; Keynes Papers, PS/5, King's College, Cambridge. The (revised) passage in the article is published in KCW/XXI:34.
16 The work referred to is Henderson (1931). Henderson was a socialist of Marxist persuasion with several popular political-economic books to his credit, each aiming at clearness of argument and minimal jargon.
17 This was altered to 'Is an individualist economy capable of providing full employment?' in mid-1935 (KCW/XIII:525), before becoming the less direct 'Concluding notes on the social philosophy towards which the General Theory might lead' in the published work.
18 Unlike Marxism, according to Keynes, which accepted the 'Classical hypothesis'.
19 In certain respects in this area, he was similar to Mill and Marshall. On Marshall's 'tendency' to socialism, see, for example, McWilliams-Tullberg (1975) and Groenewegen (1995). The necessity and difficulty of being clear on the meaning of socialism may be illustrated with reference to the discussion of Marshall's political philosophy, which in some (but far from all) respects was similar to Keynes's. If

Marshall is a 'tendential socialist', then what is the benchmark of socialism towards which his views are tendencies? What is full or 'non-tendential' socialism? If it were agreed that Marshall was a tendential socialist, then Keynes, who was significantly further down the road (especially in regard to the role of the state), must be more than a tendential socialist, unless we postulate degrees of tendencies. But this only brings us back to the central issue—the need for a general conception of socialism against which to measure tendencies and their degrees.

20 However, in a 1910 paper, 'On the principle of organic unity', Keynes introduced a modification which could be used to assist the evaluation of political and economic arrangements. While not abandoning the primacy of ethics, he argued that states of affairs were capable of possessing attributes which had a kind of ethical value or 'desirability' in themselves. We should therefore aim not only at increasing goodness but also the desirable attributes in states of affairs. Of the examples of attributes he listed—beauty, harmony, justice, tragedy, virtue, consistency and truth —some can clearly be linked to political and economic states of affairs, but Keynes did not himself explicitly make this connection in a paper focused primarily on ethics.

21 Even if the probable advantages were equal, Keynes's principle of moral risk favours cautious rather than risky action. On this principle, see O'Donnell (1989: 122 – 7).

22 The detailed case for Keynes as the author of a new kind of socialism was presented in O'Donnell (1989: chs 13, 14; 1991, 1992). Neither Skidelsky (1992) nor Moggridge (1992), for example, have shown any interest in this case.

REFERENCES

References to the Royal Economic Society's edition of *The Collected Writings of John Maynard Keynes* (Keynes 1971 – 89) take the form KCW/volume number: page.

Dalton, H. (1953) *Call Back Yesterday: Memoirs 1887 – 1931,* London: Frederick Muller.

Durbin, E. (1985) *New Jerusalems: The Labour Party and the Economics of Democratic Socialism,* London: Routledge and Kegan Paul.

Groenewegen, P. (1995) *A Soaring Eagle: Alfred Marshall 1842 – 1924,* Aldershot: Edward Elgar.

Harcourt, G.C. (1986) 'Making socialism in your own country', in C.Sardoni (ed.) (1992) *On Political Economists and Modern Political Economy: Selected Essays of G.C.Harcourt,* London and New York: Routledge, 281 – 96.

——(1996) (ed.) *Capitalism, Socialism and Post-Keynesianism,* Aldershot: Edward Elgar.

Henderson, F. (1931) *The Economic Consequences of Power Production,* London: Allen and Unwin.

Keynes, J.M. (1971 – 89) *The Collected Writings of John Maynard Keynes* (KCW), London: Macmillan.

Liberal Industrial Enquiry (1928) *Britain's Industrial Future,* London: Ernest Benn.

Marx, K. (1847) *The Communist Manifesto,* Harmondsworth: Penguin 1967.

McWilliams-Tullberg, R. (1975) 'Marshall's tendency to socialism', *History of Political Economy,* vol. 7, no. 1, 75 – 111.

Moggridge, D.E. (1992) *Maynard Keynes: An Economist's Biography,* London and New York: Routledge.

O'Donnell, R.M. (1989) *Keynes: Philosophy, Economics and Politics,* London: Macmillan.

——(1991) 'Keynes's political philosophy', in W.J.Barber (ed.) *Perspectives on the History of Economic Thought,* vol. V, Aldershot: Edward Elgar.

——(1992) 'The unwritten books and papers of J.M.Keynes's, *History of Political Economy,* vol. 24, no. 4:767 – 817.

——(1995) 'A supplementary edition of J.M.Keynes's writings: rationale, nature and significance', *History of Economics Review,* no. 23, winter.

——(1998) 'Mixed goods and social reform', *Cahiers d'Economie Politique,* no. 30 – 31, 167 – 85.

Skidelsky, R. (1983) *John Maynard Keynes: Hopes Betrayed, 1883 – 1920,* London: Macmillan.

——(1992) *John Maynard Keynes: The Economist as Saviour, 1920 – 37,* London: Macmillan.

——(1996) *Keynes,* Oxford: Oxford University Press.

9
ON SOME ARGUMENTS FOR THE RATIONALITY OF CONVENTIONAL BEHAVIOUR UNDER UNCERTAINTY
Concepts, applicability and criticisms

David Dequech

There exists an increasing body of literature that tries to show that emphasising uncertainty in a strong sense (such as Keynes') does not lead to theoretical nihilism.[1] A major part of this literature deals with the possibility of rational behaviour under uncertainty, often resorting to Keynes' references to conventions.

In what follows, I try to organise, develop and criticise the treatment of the relation between conventions and rationality under uncertainty, which has been rather obscure and confusing, particularly in what concerns the concept of convention and the arguments in favour of the compatibility between conventions and rationality. This is the first step in what I consider to be a more general and proper approach, which I pursue in another paper (Dequech 1996).

THE CONCEPT OF CONVENTION

Several definitions of convention coexist in the literature that refers to conventional behaviour under uncertainty. Since some of these definitions are interpretations of what Keynes meant by a convention, it is worth noting that Keynes did not explicitly define convention. He (1936:152 – 3; 1973c: 114, 124) can be summarised as referring, rather loosely, to first, the assumption that the current situation will continue to exist in the future unless specific indicators in the contrary direction appear,[2] and second, the resort to the average or majority opinion. The same applies to other authors who follow Keynes—e.g. Lawson (1985:916), Carabelli (1988, esp. 225, 227,237, 301), O'Donnell (1989:254, 261) and Meeks (1991:128 – 32). In my opinion, this is too vague.[3]

I have selected a few of the more explicit definitions.

1 Some writers explicitly argue for a broad concept of convention. Darity and Horn (1993:28, 31), interpreting Keynes, mean by a convention a rule of thumb and state that 'there are many types of rule of thumb'. The projection of the existing situation and the reliance on the opinion of others would be examples of rules of thumb.[4]

2 Also inspired by Keynes, another conceptualisation, hinted at by Possas (1990:11 – 12) and developed here more explicitly, identifies convention primarily as conformity with the average opinion. The projection of the present and the recent past into the future is considered as one—even if the most important—of the possible conventional (that is, average) opinions. The projection of the existing situation would be, then, one of the specific expectational patterns that people can adopt.

3 On my own definition, convention is, first of all, a *collective* rule of thumb, or, equivalently, a socially shared standard of behaviour. This is different from definition 2 in that the average opinion can be the result of many different individual types of behaviour, whereas the idea of a collective rule of thumb or standard of behaviour suggests that many people follow a similar rule or standard. The two definitions are almost equivalent when the dispersion around the average opinion is very low. Then, the specific pattern of expectations that represents the average opinion would be the one underlying the specific rule of thumb that is followed. As a socially shared and regular standard of behaviour, a convention on this definition is similar to what other people and I call an institution or a type of institution. On my definition, a convention has two other characteristics: first, it is followed at least in part because other people are following it; second, it is to some degree arbitrary.

4 For Lewis (1969:58), a convention is also a collective standard of behaviour, but his definition is more restrictive than mine, for it requires everyone to conform to the convention when there is an expectation that others will do the same. Based on Lewis, Bromley (1989:41 – 2) and Runde (1994:248) adopt a similar definition (Bromley 1989:79n, notes that Lewis' definition is not always so strict)—see also Schotter (1981) and Sugden (1986).

The rule-of-thumb definition proposed by Darity and Horn (1993) applies, as in Simon, to an individual procedure which is not necessarily associated with a collective pattern of action, although the resort to the opinion of others is seen as one among the several possible individual rules of thumb. The definitions of convention as compliance with the average opinion and as a collective rule of thumb are closer to the common way ofseeing a convention as something shared with others (see also Lawson 1991a:221; McKenna and Zannoni 1993:402; Dow 1995:119).

These conceptual differences are very relevant for an analysis of the relation between convention and rationality, to which I now turn.

A CRITICAL EXAMINATION OF SOME ARGUMENTS FOR THE RATIONALITY OF CONVENTIONAL BEHAVIOUR

Before anything else, I should emphasise the obvious fact that, if a convention is seen as a response to uncertainty, the study of a possible compatibility between convention and rationality cannot adopt a concept of rationality (such as in mainstream neoclassical economics) that is indissolubly linked to the assumption that Keynes' uncertainty, if it exists, is irrelevant. Otherwise, the study is over as soon as it starts and the possible compatibility is denied.

Different interpretations are presented in the following paragraphs which share a belief in the compatibility between convention and rationality. In this sense, they are contrary to the position taken by Shackle (1967:228), who treats the resort to conventions as a substitute for reason, the latter being incapable of operating for lack of data.

Several authors working with a strong concept of uncertainty broaden the notion of rationality by associating it with the existence of 'good reasons' or with people 'doing the best they can under the circumstances'—see Lawson (1985: 918; 1993:194), Harcourt (1987:237), Hamouda and Harcourt (1988:214), O'Donnell (1990:258, 264), Runde (1991:134), Lavoie (1992:51), Langlois (1986:230). Without further specification, this is too generic. In particular, this could be compatible with an extremely subjectivist view. To avoid extreme subjectivism, rationality is defined here as something that requires knowledge, in a somewhat objective sense. Uncertainty does not imply complete ignorance and therefore allows for a broader notion of knowledge and rationality than the neoclassical. Irrationality is that which is contrary to rationality. I overcome the dichotomy between rationality and irrationality by introducing a third possibility: arationality is not based on, nor contrary to, knowledge. Hence, under uncertainty there is necessarily some space left for arationality, when knowledge is not enough to determine what is a rational and, consequently, what is an irrational course of action. In addition to being based on knowledge, rational behaviour is also defined here as the most adequate, in the light of the available knowledge, to the achievement of some end, even if it turns out to be unsuccessful ex post. This criterion is particularly useful when analysing the behaviour of capitalists, decision-makers who, at leastfor some analytical purposes, can be assumed to be self-interested and to have pecuniary gain as their ultimate objective.

Another preliminary remark is necessary. Following Lewis and others, game theory has discussed convention (or institution, as it is sometimes called), defining it as a solution to a coordination game with multiple equilibria (for uncomplicated, recent discussions see Young 1996 and Lecq 1996). I will not deal with the relation between convention and rationality in game theory, for the latter traditionally abstracts from uncertainty in the strong sense adopted here.[5]

There are two basic ways of approaching the relation between convention and rationality under uncertainty so that compatibility between them is possible. In the first and prevalent one, the relation between convention and rationality has been constructed in a restrictive manner, which centres on the rationality of *following* a convention. See Kregel (1987:525), Harcourt (1987:241), Feltz and Hoogduin (1988:117), Rotheim (1993), Rogers (1997: 330). More specific references to several other authors will be made shortly. The second approach, which I suggest in Dequech (1996), is less restrictive. Here I will limit myself to a discussion of the first approach.[6]

Different arguments may be given to try to show that it is rational to follow a convention, depending, among other things, on how the latter is defined. In what follows I present some of these arguments, try to improve and/or generalise them, but also ultimately criticise them. It should be noted from the start that my own position is that following a convention is not necessarily the best option to take. I prefer the second, less restrictive, approach to the first. In this sense, I will be surveying and even trying to improve some arguments which I eventually consider insufficient, at least in the light of my own definition of rationality. The approach that asserts the rationality of following a convention under uncertainty is not completely successful. A particularly important type of action which this first approach does not adequately deal with is innovative behaviour, as in the case of the Schumpeterian entrepreneur. Even if one improves the definition of convention and the arguments for following a conventional behaviour, as I will try to do, one cannot argue that it is best for a decision-maker to follow a convention. I should also note that my critical assessment of the first approach is completed only in the sequel to this paper, with my proposal of an alternative. This alternative goes beyond the literature discussed here in two aspects. First, it expands the discussion of the rationality of conventional behaviour by explicitly connecting conventions, tacit knowledge and rationality. Second, it incorporates the possibility of unconventional behaviour as something compatible with rationality.

Workability

The rationalisation of conventional behaviour in terms of 'workability' is most clear in Darity and Horn (1993:29), who define convention as an individual rule of thumb: Adherence to a rule of thumb appears—at least for a time—to make affairs manageable'. 'There is no alternative that is, typically, more workable' (*ibid:* 32). 'Workable' seems to mean here not too complicated.

The argument of workability applies both to financial and product markets.[7] Rules of thumb—be they individual or collective—save the time involved in analysing a large amount of information.[8]

Now one can ask which specific rule of thumb the individual will choose and why. The problem is that being twice as bold as the average opinion is also

workable, and so is doing nothing and preferring liquidity. Some elements of this discussion appear in the next sections. Moreover, uncertainty is not merely complexity. Thus, the workability argument is not sufficient to rationalise conventional behaviour under uncertainty.

Convention as induction

Another argument intended to show the rationality of conventional behaviour appears in the writings of authors who propose to interpret *The General Theory* in the light of Keynes' major philosophical work, his *Treatise on Probability* (TP). According to that interpretation, it is rational to use induction as a guide for action under uncertainty. O'Donnell (1989) and Gay Meeks (1991) are among the main defenders of such view. For example, O'Donnell (1989:250) argues that, in Keynes' theory of expectations, 'the primary notion is that expectations are generally based on *induction*' (original emphasis). O'Donnell clearly implies that induction is what Keynes has in mind when he speaks of projecting the existing situation into the future (see also Meeks 1991).[9]

It is clear that O'Donnell's interpretation is centred on the projection of the existing situation. To the extent that in this argument the projection of the existing situation can be associated with a convention,[10] then conventional behaviour would be seen as rational behaviour. However, I draw a distinction between the rationality of *belief and* the rationality of *behaviour*. Rational belief is a necessary but insufficient condition for rational deliberate behaviour. The rationality of deliberate behaviour depends on the reliability of rational belief as a guide to conduct. This distinction is useful when discussing the TP. In the TP, probability is a rational degree of belief. Probability as applied to conduct has to be supplemented by weight (TP: ch. 26). Even if some kind of probable knowledge (in the TP sense) were possible under uncertainty,[11] there would still exist doubt as to the reliability of such knowledge as a guide to action (O'Donnell himself acknowledgesthis reliability problem, as I discuss below). Similarly, in Keynes' later writings, the notion of confidence reflects the doubts as to the reliability of expectations. The corresponding distinction between expectations and behaviour should also be made. Therefore, conventional behaviour cannot, under uncertainty, be completely based on inductive or probable knowledge. Consequently, it cannot be completely rationalised on this basis. Moreover, if the projection of the current situation into the future were completely based on knowledge, it would not be arbitrary. Then, if a convention is to some degree arbitrary, by definition, how could that projection be called a convention? Be it conventional or not, behaviour under uncertainty has to depend, at least in part, on animal spirits and the like (contrast this with Rogers 1997:330). This is valid also for short-period decisions, which do involve uncertainty, contrary to what seems to be O'Donnell's view.

Furthermore, some people may intentionally break with the current situation through their own innovative actions, as I discuss in Dequech (1996). The Schumpeterian entrepreneur does not project the existing situation into the future, but instead creates a new future.

Convention as a self-fulfilling prophecy

Hinting at a definition of convention as conformity with the average opinion, Possas (1990:11 – 12) argues that it may be rational to follow the convention because the average opinion may indicate 'a convergence, and, as such, a probable market trend' of the variables considered. The use of the average opinion as such an indicator has 'at least two preconditions…. The first consists in the possibility of identifying, and progressively turning into conventional use, a particular pattern of expectations'; the second is 'a low dispersion of opinions around the average'. The first precondition is necessary for the formation of the convention, which obviously is, in this perspective, a specific average opinion; the second is required so that the convention is really representative of a likely market trend.

In this argument, as I understand it, the average opinion represents the best expectation about future results in a particular market, and to follow the convention is to try to make the right guess. This argument logically depends on the potentially self-fulfilling character of the average opinion. Keynes (1936) could perhaps be seen as applying this to the expectations about the interest rate (see also Freedman 1995:91).[12] At any rate, Possas (1990:23n) believes that the average opinion can represent a likely trend of product markets, under some stable conditions. This is contrary to O'Donnell's (1989:261) belief that the average opinion would be of most significance for those operating in financial markets (also, although less emphatically, Meeks 1991:132 – 3, 136).

Concepts of convention that refer to an eminently individual behaviour,with no necessary relation to the opinion of others, cannot give rise to this kind of rationalisation of conventional behaviour.

This argument is critically examined in a later section.

Conventional behaviour as defensive behaviour

One may try different ways to defend the rationality of following a convention by saying that this is defensive behaviour in conditions of uncertainty.

One of them is provided by Lawson (199la). He centres his analysis of conventional behaviour on the projection of the present and recent past, which is for him one example of a convention. He confines this argument of defensive behaviour (and interprets Keynes as doing the same) to the 'investment' decision and, even more restrictively, to the decision of 'investing' in financial markets.[13] Holding to the projective convention allows investment to be made safe (*ibid.:*

194, 207). Lawson supports this argument with Keynes' (1936:152) reference to 'organised investment markets'[14] where 'investments' are made liquid for the individual, although not for the community as a whole, as long as the individual can 'rely on the maintenance of the convention'. Lawson (1991a:208, original emphasis) states that

> it is essential for the argument that the convention in question be viewed not merely as an individual rule of thumb…but as a procedure of action that a community of investors simultaneously draws upon…. [A]n individual investor must be able to *rely* upon it being (through the actions of others) *maintained*. Its maintenance…can facilitate a degree of continuity and stability in market evaluations and thus a degree of security

for the individual. Thus the mere projection of the current situation by a single individual is not necessarily seen as rational. Its rationality depends on other people also following the same convention. Consequently, there is a connection with the opinion of the majority, even though the concept of convention is not explicit.

Lawson's (1991a) argument applies only 'if there exist organised investment markets' (Keynes 1936:152) such as the Stock Exchange. In this context, it seems that the need for other people to keep holding to the convention of projecting the present and the recent past is important in the sense that this is what guarantees for an individual the 'opportunity to revise his judgement and change his investment, before there has been time for much to happen' (Keynes 1936:152, quoted by Lawson 1991a:194). Thus an important part of this argument is the possibility that the individual disagrees with the majority and decides to abandon the convention, which they can only do because there is an organised market. This is why Keynes refers to 'investment' being made liquid for the individual.

Those willing to rationalise conventional behaviour as defensive behaviour could use the concept of convention as a collective rule of thumb or, what can be similar in some circumstances, as conformity with the average opinion. The argument with this concept would be that conventional behaviour is a rational defence against uncertainty because it represents for the individual a potentially successful attempt to preserve their position relative to other people in the relevant market, since the individual will be behaving in a manner similar to these other people, on average.[15] This would be a more specific way of arguing that 'the safest course of action may simply be to follow the crowd' (Hamouda and Smithin 1988b:281) or that 'there are less chances of getting burned when one is following the crowd' (Lavoie 1992:57). Again, for the average opinion to be representative, the dispersion around it has to be low.

Possas (1987:133 – 4) provides this type of argument in his analysis of conventional behaviour in the investment decision (in the sense of the decision to

buy capital goods). For him (as for Shackle 1967:133), Keynes gave too much attention to the connection between (physical) investment and speculation in financial markets, thus overemphasising the forecasts expressed in these markets and neglecting the conditions of competition in the product market. This interpretation is subject to qualifications, for example regarding Possas' belief that Keynes restricted the use of his idea of convention to financial markets. Nevertheless, Possas (1987:133n) is right in pointing out the importance, for an individual's decision to invest in a product market, of the opinion of competitors in that market.

This argument is relatively weakened, however, if presented without an explicit concept of convention, as by Possas (1987), who follows Keynes' 1937 QJE article (see also Lawson 1991b:245 – 6). In my opinion, once the definition of convention as a collective rule of thumb or as compliance with the average opinion is adopted,[16] the argument of the defensive character of conventional behaviour gains precision and a more visible generality, being applicable not only to the investment decision in the product markets but also to the production decision in these markets[17] and even to financial markets[18] and possibly the labour market.[19] Again, this is contrary to O'Donnell's (1989) and Meeks' (1991) belief that the average opinion would be of most significance for decisions in financial markets.

Being based on the attempt to maintain the relative position of the individual in a market, and different from Lawson's (199la) justification of the defensive character of conventional behaviour, which centred on the convention of projecting the present, this argument does not depend on the specific pattern of conventional expectations, that is, on what the average opinion expects the future to be. When this specific pattern is the projection of the current situation, regardless of whether it is more difficult to rationalise this projection by a single individual isolatedly, it would be argued in this case that it is rational to project the existing situation because the other decision-makers, on average, are doing the same.[20] More generally, it would be rational to do whatever the other decision-makers, on average, are doing.

Neither does conventional behaviour as defensive behaviour depend on the convention tending to be a self-fulfilling prophecy. Even if the collective opinion turns out to be contradicted ex post, following it would help the preservation of the relative position by the individual.[21]

Therefore, at least as far as product markets are concerned, the defensive character of conventional behaviour would provide a more general rationalisation for this behaviour than the argument of the self-fulfilling prophecy, discussed above. The latter can be seen as complementary to the former, reinforcing it in some particular cases which depend on the market in question.[22]

The argument in favour of the rationality of defensively following the average opinion or a collective rule of thumb, would be more general also in the sense of being applicable, in what concerns products markets, both to the investment and

the production decisions in these markets. Therefore this argument, associated with the distinction between the convention and the specific pattern of conventional expectations, would imply the necessity of qualifying O'Donnell's (1989:253) position, according to which 'it is in relation to the formation of LTE [long-term expectations] that rational behaviour becomes a more complicated, fluid and precarious affair'. His observation does not apply to the collective rule of thumb or to conformity with the average opinion, although it may be valid for the specific pattern of conventional expectations, such as the projection of the current situation, which I will discuss shortly.

Behind the difference between the defensive argument and O'Donnell's in this regard lies the fact that he primarily links rationality to induction and associates the projection of the present and the recent past with induction. As shown above, for O'Donnell (1989:250), induction is the 'core notion' in Keynes' theory of expectations. O'Donnell (1989:252) argues that to this 'core notion' of induction 'Keynes attached two further ideas'. The first is that only knowledge of the present and the recent past, as opposed to the distant past, 'will commonly be used in the inductive projection'. The second 'is that the reliability of induction as a guide to the economic future decays rapidly when expectations are pushed beyond the immediate horizon'. This is why for O'Donnell the rationality of behaviour based on long-term expectations is a more complicated affair. Meeks (1991:153 – 4) argues similarly.

With the definition of convention based on conformity to the average opinion or on a collective rule of thumb, the other arguments do not depend on the projection of the existing situation necessarily being the specific pattern of expectations chosen, and apply equally to short- and long-term expectations.

As in the case of the self-fulfilling-prophecy argument, I present my objection to the defensive argument in the following section.

The problems with following the average opinion or a social rule

My first criticism applies to both arguments presented in the previous two sections, that is, to the attempt to rationalise conventional behaviour as the pursuit of a self-fulfilling prophecy and/or as defensive behaviour. The trouble emerges from the difficulty an individual may find in ascertaining the average opinion in the market they operate in. There may exist some information in this regard, but other information is missing. Uncertainty involves not knowing what others will do.[23] This problem has to do with the contribution of the interdependence between decision-makers to the uncertainty faced by them, and also with the possibility of creative behaviour on the part of other decision-makers. Even if decision-makers have been following a convention, some of them may lose their confidence in it and abandon it. Others may intentionally adopt an unconventional behaviour. People may have knowledge of conventions,

as Lawson (1985:916 – 17) and Darity and Horn (1993:30) claim, but they also must have doubts about the continuity of conventions and the reliability of such knowledge as a guide to action. My point is thus basically the same I made when commenting above on convention as induction. By the way, uncertainty implies also that the resort to the opinion of others cannot, in general terms, be justified on the basis, suggested by Keynes (1937:114) that 'the rest of the world' is 'perhaps better informed' than a particular individual. It is true that, if uncertainty does not imply complete ignorance, some people may be better informed than others. However, if decision-makers are aware of uncertainty, they know that other people also lack some knowledge and therefore are not totally reliable. Moreover, a decision-maker has also to consider that other people may be *less* informed than they are.

The second criticism has also to do with uncertainty and is directed more specifically to the argument that conventional behaviour is rational because it is defensive. The definition of rationality adopted here relates rationality to the adequacy of behaviour to some end. It is possible to characterise defence as an end. Lawson (1991a:215), for example, argues that 'in defending the adoption of conventions by investors, Keynes effectively shifts the objective to investment safety'. 'To fare no worse, generally, than others' (Lawson 1991b:245) would be a shorter-term objective. The point made in the previous paragraph raises problems even if the rationality of behaviour is assessed in relation to this objective. If there are doubts as to what others will do, there are doubts as to how one will fare in comparison with others. Uncertainty implies such doubts, even if to some degree the maintenance of the convention can be relied upon. However, there is still another problem. A non-ultimate end can be seen as a means to an ultimate end. Lawson (1991a:214) himself refers to a general motive 'which provides a criterion against which other "ends" can themselves be assessed'. What happens then if this general end is the pursuit of pecuniary gain, as suggested above?[24]Even if one knew with certainty what the others are going to do, one would still have doubts as to whether following the others would be adequate to the pursuit of this ultimate end.[25] There are competing alternatives. One of them is to adopt the position that is really defensive: to prefer liquidity. Another option would be to intentionally disobey the convention, as the Schumpeterian entrepreneur does.

I should also note that there is another type of argument that might be used to rationalise conventional behaviour. More specifically, this argument applies to the following of social norms, which, as I understand them, are conventions with normative content. Norm-guided behaviour can sometimes be justified by the idea that people who follow a social norm are rational if they are pursuing a particular objective: to avoid the criticism of others, or even ostracism. If this is the ultimate objective, conventional, norm-guided behaviour will achieve it almost by definition, although the problem of not knowing what others will do and thus not knowing how to compare ourselves to others may also be relevant here.

A particular example of this in economic decisions is perhaps the one about which Keynes (1936:157 – 8) writes: 'Worldly wisdom teaches us that it is better for reputation to fail conventionally than to succeed unconventionally' (see also Lavoie 1992:57). Lawson (1991a:217 – 18) applies this reasoning to the situation of managers of 'investment' funds in organised 'investment' markets. Davidson (1994:237 – 8) uses it when discussing the case of international bankers and comptrollers of multinational corporations operating in exchange rate markets. The argument can be extended to managers of large corporations dealing in product markets within a national economy (for a formal discussion, in varied contexts, see Scharfstein and Stein 1990). However, in all these cases the objective is assumed to be to protect the person's reputation. A procedure which is rational given this end will not necessarily be so if we consider a different objective, such as the pursuit of pecuniary gain. The reputation argument depends on specific even if quite common institutional settings. Moreover, the combination of reputation and profit considerations is quite complex. Succeeding unconventionally is not necessarily harmful to a person's reputation (see also Choi 1993), but under uncertainty there are serious doubts about the chances of success.

A further comment on the projection of the existing situation

Keynes (1936:249; 1973b:137) pointed out that the economic system 'is not violently unstable' and that one 'must not confuse instability with uncertainty'.[26] Following these warnings, Davidson (1978:385) argues that

> the economic system is *potentially* very unstable. Recognising the mercurial possibility of the economic system, man has, over time,devised certain institutions and rules of the game, which, as long as they are operational, avoid such catastrophes by providing a foundation for a *conventionality* of belief in the stability of the system and hence in the quasi-stability of the state of expectations.
>
> (emphasis added)

Among these institutions, Davidson (1978:388 – 9) highlights, first of all, contracts and especially the money-wage contract, crucial for the 'conventionality of belief in price stability.[27] He (1994:49 – 50) also emphasises the existence of a 'market maker' in national and international financial markets, which provides a considerable degree of orderliness in these markets. Kregel (1980:46; 1981:69 – 70) refers not only to wage and debt contracts but also to supply and trading agreements as 'uncertainty-reducing institutions'.

Furthermore, along with these institutions, conventions themselves play a decisive stabilising role.[28] As a particular case, the convention of projecting the

present and the recent past into the future contributes to this, as Keynes himself argued,[29] and therefore reinforces itself. The point is more general: institutions, conventions, social practices, etc., through their influence on people's behaviour, bring order to economic life. This point, forcefully made by such institutionalists as, for example, Neale (1987), Hodgson (1988) and Fusfeld (1989), is acknowledged by Carvalho (1983 – 4) and Lawson (1985).

These factors make the projection of the present and the recent past a pattern of expectations particularly capable of being adopted by many different people, and thus of becoming a convention. They justify Keynes' emphasis on this projection. Especially for short-term expectations, this convention can be said to be *partly* based on some type of knowledge (even if one does not call it inductive or probable knowledge).

It is more difficult for this projection to become the specific expectational pattern adopted by the average opinion in the case of long-term expectations, which guide the decision to purchase capital goods. In this specific sense, O'Donnell (1989:253) correctly highlights the difference between short- and long-term expectations. For example, contracts with the same, very long duration as capital goods are not so common. Moreover, the more distant the future, more time is allowed for technical innovations or important cultural and political changes to occur, for example.

The projection of the existing situation might even be seen by the average opinion as the best long-term expectation to form in the absence of a calculable basis. People can have some confidence in this convention, and particularly they will have more confidence in it in a country with more stable institutions and social practices, *ceteris paribus*. However, this confidence tends to be considerably less than in the case where the projection of the current situation is the conventional short-term expectation. This is what makes investment expenditures especially liable to sudden changes.

CONCLUDING REMARKS

This paper has tried to organise and develop the discussion of what is meant by convention and why and where it would be rational to follow a convention under uncertainty. It has also ultimately pointed out problems in the attempts at rationalising conventional behaviour. Given the definition of rationality adopted here, according to which rationality requires that deliberate behaviour be based on knowledge and directed at some specified ultimate end (for example, pecuniary gain), all these problems originate from uncertainty. Uncertainty poses different specific difficulties for the different arguments, but in general it is impossible to completely rationalise behaviour under uncertainty, since the latter implies some ignorance, at least as to the reliability of the possible knowledge as a guide for behaviour. This applies for whatever knowledge of conventions people may have. For all the arguments, except the workability one, there is also

the uncertainty as to whether a convention that has been followed for some time will be maintained.

Rationality as something that requires knowledge can be gradable (as can knowledge and its counterpart, uncertainty). In this sense, conventional behaviour can be seen as rational, but only partially, in the same way as people have only partial knowledge under uncertainty (such as the knowledge about uncertainty itself, as well as about institutions and conventions).

This paper has not implied, however, that convention is incompatible with rationality. I propose in Dequech (1996) an approach in which, contrary to the arguments discussed above, the compatibility between convention and rationality is not restricted to the case in which people follow a convention.

ACKNOWLEDGEMENTS

I am grateful to Paul Davidson, Alessandro Vercelli and the participants of the Post-Keynesian Economics Workshop, Cambridge, February 1996, for comments on earlier drafts of this chapter. Above all, I would like to express my gratitude to Geoff Harcourt and my admiration for his very rare combination of scholarship, kindness and joie de vivre. Financial support from CNPq/Brazil is thankfully acknowledged.

NOTES

1 The charge of nihilism was made by Alan Coddington (e.g. 1982), to whom part of that literature responds directly.

2 This is different both from the Rational Expectations Hypothesis and the Adaptive Expectations Hypothesis.

3 Excessive broadness rather than vagueness is the potential theoretical problem with Davis' (1994a:166; also 1994b) conceptualisation of Keynes' convention. Convention would be a structure of different individual expectations 'that bears a complex relationship to average opinion as its central reference'. This allows procedures that differ from a certain standard to be considered a convention. In this case, it would be difficult to define what unconventional behaviour is. Consequently, convention itself becomes too broadly defined. In contrast, the theoretical interest of the definition extracted from Dutt and Amadeo's (1990: 105 – 6) reading of Keynes is affected by their primary association of convention with the projection of the existing situation. Relying on other people would not be a convention. This seems to be too restrictive.

4 Although not so explicitly, Lavoie (1992:55 – 8), combining Keynes and Simon, takes a very similar position. For Driver and Moreton (1992:73 – 4), likewise, the notion of convention encompasses 'a broader range of possibilities than merely maintaining previous behaviour', but they fail to be more specific than this.

5 See Shackle (1972:422 – 6) and, regarding the Bayesian foundations of game theory, Binmore (1987:210 – 11). Other related criticisms of this literature on conventions/institutions in game theory are made by Mirowski (1986). I should

note, however, that game theory has begun to incorporate recent developments in decision theory which go beyond the standard, weak notion of uncertainty (or risk) prevalent in mainstream economics (e.g. Dow and Werlang 1994; Lo 1996; see also Epstein and Wang 1996). It is still to be better assessed whether these developments really convey the strong notion of uncertainty involved in the most relevant economic decisions, as in Keynes' sense. Even less explored so far is the impact of these developments for the discussion of rationality and conventions in a game-theoretical application. The discussion of rationality in game theory is already very complex, to say the least, without introducing issues related to strong uncertainty.

6 Before doing this, I should note that there are those such as Fitzgibbons (1988) and Winslow (1989:1180), who, specifically dealing with Keynes, identify an *incompatibility* between convention and rationality. At least in part, this involves Keynes' position on ethics (and G.E.Moore's influence on it), which lies beyond the confines of this paper. On this, see the rebuttal of Fitzgibbons' claim by Lawson (1991a:204 – 8; 1993). I expect to provide below several arguments for qualifying Fitzgibbons and Winslow's interpretation without having to discuss ethics. I will avoid discussing the rationality or otherwise of pursuing money or pursuing the good, issues that appear in Winslow (1989, 1992).

7 Even if he did not (at least not explicitly) define convention as a rule of thumb, Keynes (1936:51) can be interpreted as using 'workability' as part of his justification for the convention of projecting the present when forming short-term expectations: 'It would be too complicated to work out the expectations *de novo* whenever a productive process was being started'.

8 This is clearly close to Herbert Simon's (1955, 1959) work on bounded rationality. The rationalisation of rule-following in terms of its workability for an individual with a limited capability to deal with a vast and complex amount of information also appears in other traditions of economic thought. Rutherford (1994:59 – 60) and Perlman (1986:272) point out similar ideas by old institutionalists such as Commons, Mitchell and J.M.Clark. Hayek should also be mentioned here, at least in Vanberg's (1993:181 – 2) interpretation (but see Vromen 1994:216). A more recent author to emphasise the rationality of rule-following people with what he calls a competence-difficulty gap is Heiner (1983).

9 On Keynes and induction, see also Brown-Collier and Bausor (1988), Hamouda and Smithin (1988b) and Carabelli (1988).

10 It can indeed, because these authors do not explicitly define convention. Following Keynes' 1937 Galton Lecture (Keynes 1973c:124), O'Donnell (1989: 254) refers to the 'inductive process described above' as a convention. On another occasion (*ibid:* 261), he follows Keynes' 1937 QJE article association between conventional judgement and imitation of the majority. The same applies to Meeks (1991). Another thing to note regarding this line of argument is that a convention does not need to have a social character. It can be merely individual.

11 Lawson (1991a:200; 1995:94 – 5) should also be mentioned for arguing that there are epistemic reasons for projecting the present and the recent past, even if not referring to induction (see also Dow and Dow 1985:52; Runde 1991:136 – 7). Lawson (1995) also adds a psychological motivation, which I discuss in Dequech (1996). At any rate, I associate his position with another argument, discussed below.

12 Referring to the rate of interest as a conventional phenomenon, Keynes (1936: 203) writes: 'its actual value is largely governed by the prevailing view as to what its value is expected to be. *Any* level of interest which is accepted with sufficient conviction as *likely* to be durable *will* be durable' (original emphasis). The use of the notion of convention in this case is complicated, however, for Keynes (1936: 172) also mentions the need for divergent opinions in such markets.

13 Keynes, in his analysis of the stock exchange (1936:152 – 3), refers to 'investment' and 'investors' also in a less strict sense than the one that applies (as in Keynes 1936:62; 1937:118) only to the direct purchase of capital goods.

14 An organised market is one where there exists what Davidson (1994:49 – 50) calls a 'market maker'. I refer below to the relation between the existence of this type of institution and the convention of projecting the current situation.

15 Therefore this behaviour is conventional in two complementary senses: first, it is, by definition, something socially shared; second, it is conservative, not bold.

16 This definition, as argued above, makes explicit and develops the connections, intimated by Possas himself in another paper (1990), between convention, conformity with the average opinion and projection of the current situation. In his 1990 paper Possas does not refer to this defensive preservation of the relative position, restricting himself to the argument of the self-fulfilling prophecy discussed in the preceding section.

17 This runs counter to the position of Littleboy (1991:278).

18 For example, this argument would support Darity and Horn's (1993:29 – 30) analysis of herd behaviour, in which they extend Keynes' discussion to another segment of the financial markets, namely bank lending. As seen in the previous section, for them herding is one of the possible rules of thumb that people may follow.

19 This could provide a rationale for the concern of workers with their relative wages, as argued by Keynes (1936: ch. 2), in a situation where they cannot bargain in terms of real wages and there is uncertainty as to what the prices of wage-goods and thus real wages will be. Workers trying to preserve their relative wages could then be seen as following the crowd.

20 This would be more clearly applicable to a hypothetical situation in which:

(a) the projection of the current situation of a market is the specific pattern of expectations adopted, on average, by people in that market;
(b) an individual decision-maker, knowing this, follows the average opinion by making the projection.

Reality, however, may be quite different, as I discuss below.

21 It should be noted that a different type of argument applies to the case in which people want to protect their reputations. I come back to this below.

22 It should be noted that both arguments, and perhaps especially the defensive one, involve an evaluation (not totally based on calculus) of the consequences of conventional behaviour. They are not necessarily in conflict with Keynes' 1938 letter to Townshend (Keynes 1979:294) where he mentions motives 'which are not "rational" in the sense of being concerned with the evaluation of consequences'. The compatibility lies in that Keynes may be interpreted as referring to a *calculus-*

based evaluation, since none of the alternatives 'is demonstrably more "rational" than the others, in the sense that we can arrange in order of merit the sum aggregate of the benefits obtainable from the complete consequences of each'.

23 Contrast this with Darity and Horn (1993). According to Shackle (1970:102, *apud* Kregel 1990:90), 'a great part of business efforts is directed to defeating the efforts of one's rivals to know what one is doing and is going to do'. Keynes himself wrote (1971:292) that entrepreneurs 'take pains to conceal [their decisions] from one another'.

24 Lawson (1991a:191 – 2, 203) shows that in Keynes' case the pursuit of money is accepted merely as a temporary means for the truly ultimate end, which is the search of 'good'. I do not address ethical issues here, nor do I discuss whether the profit motive is irrational or not (see Winslow 1992:104 – 5).

25 The same criticisms apply to Ewijk's (1991:62) explanation of rule-following in terms of the aim of continuity or survival.

26 In the same vein: 'the state of long-term expectation is often steady' (Keynes 1936: 162).

27 Townshend (1937) was an early interpreter of Keynes to emphasise this. See also Shackle (1967:228, 247), Carvalho (1992:28, 100 – 102) and Rotheim (1993).

28 See, for example, Vickers (1979 – 80:244 – 5; 1994:157 – 9), Harcourt (1981:261), Harcourt and O'Shaughnessy (1985:8 – 9), Earl and Kay (1985:43), Hamouda and Smithin (1988a:163), Carabelli (1988:220), Littleboy (1991:30), Lavoie (1992:11), Darity and Horn (1993:30), Rotheim (1993:199), Crotty (1994).

29 At least in the stock market context, 'the above conventional method of calculation will be compatible with a considerable degree of continuity and stability in our affairs' (Keynes 1936:152). I believe the argument is valid for many other markets.

REFERENCES

Binmore, K. (1987) 'Modeling rational players I', *Economics and Philosophy,* 3, 2, October: 179 – 214.

Bromley, D. (1989) *Economic Interests and Institutions,* Oxford: Blackwell.

Brown-Collier, E. and Bausor, R. (1988) 'The epistemological foundations of *The General Theory', Scottish Journal of Political Economy,* 35, 3, August: 227 – 41.

Carabelli, A. (1988) *On Keynes' Method,* London: Macmillan.

Carvalho, F. (1983 – 4) 'On the concept of time in Shacklean and Sraffian economies', *Journal of Post-Keynesian Economics,* VI, 2, winter: 265 – 80.

——(1992) *Mr Keynes and the Post-Keynesians,* Aldershot: Edward Elgar.

Choi, Y.B. (1993) *Paradigms and Conventions,* Ann Arbor MI: University of Michigan Press.

Coddington, A. (1982) 'Deficient foresight: a troublesome theme in Keynesian economies', *American Economic Review,* 72, 3, June: 480 – 7.

Crotty, J. (1994) 'Are Keynesian uncertainty and macro theory compatible? Conventional decision making, institutional structures, and conditional stability in Keynesian macromodels', in G.Dymski. and R.Pollin (eds) (1994) *New Perspectives in Monetary Macroeconomics,* Ann Arbor MI: University of Michigan Press, 105 – 39.

Darity, W. and Horn, B. (1993) 'Rational expectations, rational belief, and Keynes' *General Theory*', *Research in the History of Economic Thought and Methodology,* 11:17 – 47.

Davidson, P. (1978) *Money and the Real World,* 2nd edn, London: Macmillan.

——(1994) *Post-Keynesian Macroeconomic Theory,* Aldershot: Edward Elgar.

Davis, J.B. (1994a) 'The locus of Keynes' philosophical thinking in *The General Theory:* the concept of convention', *Perspectives on the History of Economic Thought,* vol. X, Aldershot: Edward Elgar.

——(1994b) *Keynes' philosophical development,* Cambridge: Cambridge University Press.

Dequech, D. (1996) 'Towards a general approach to conventions and rationality under uncertainty', University of Cambridge, unpublished.

Dow, A. and Dow, S. (1985) 'Animal spirits and rationality', in T.Lawson and H. Pesaran (eds) (1985) *Keynes' Economics: Methodological Issues,* Armonk NY:M. E.Sharpe, 46 – 65.

Dow, J. and Werlang, S. (1994) 'Nash equilibrium under Knightian uncertainty: breaking down backward induction', *Journal of Economic Theory,* 64:1343 – 73.

Dow, S. (1995) 'Uncertainty about uncertainty', in S.Dow, and J.Hillard (eds) (1995) *Keynes, Knowledge and Uncertainty,* Aldershot: Edward Elgar: 117 – 27.

Driver, C. and Moreton, D. (1992) *Investment, Expectations, and Uncertainty,* Oxford: Blackwell.

Dutt, A.K. and Amadeo, E. (1990) *Keynes' Third Alternative? The Neo-Ricardian Keynesians and the Post-Keynesians,* Aldershot: Edward Elgar.

Earl, P. and Kay, N. (1985) 'How economists can accept Shackle's critique of economic doctrines without arguing themselves out of their jobs', *Journal of Economic Studies,* 12, 1:34 – 48.

Epstein, L. and Wang, T. (1996) '"Beliefs about beliefs" without probabilities', *Econometrica,* 64, 6:1343 – 73.

Ewijk, C. (1991) 'Micro-foundations or macro-dynamics?', in W.Adriaansen and J. Linden (eds) (1991) *Post-Keynesian Thought in Perspective,* Wolters-Noordhoff: 57 – 72.

Feltz, W. and Hoogduin, L. (1988) 'Rational formation of expectations: Keynesian uncertainty and Davidson's (non)-ergodicity-criterium', *Metroeconomica,* XXXIX, 2, June: 105 – 19.

Fitzgibbons, A. (1988) *Keynes' Vision: A New Political Economy,* Oxford: Clarendon Press.

Freedman, C. (1995) 'Animal spirits in his soup: a look at the methodology and rhetoric of *The General Theory*', *Journal of Economic Methodology,* 2, 1:79 – 104.

Fusfeld, D. (1989) 'Toward a revision of the economic theory of individual behavior', *Journal of Economic Issues,* XXIII, 2, June: 357 – 66.

Hamouda, O. and Smithin, J. (1988a) 'Some remarks on "Uncertainty and economic analysis"', *Economic Journal,* 98, March: 159 – 64.

——(1988b) 'Rational behavior with deficient foresight', *Eastern Economic Journal,* XIV, 3, July-Sept: 277 – 85.

Harcourt, G.C. (1981) 'Marshall, Sraffa and Keynes: incompatible bedfellows?', *Eastern Economic Journal,* 7. Reprinted in Harcourt, G.C. (1992) *On Political Economists and Modern Political Economy,* ed. Claudio Sardoni, London: Routledge, 250 – 64

——(1987) 'The legacy of Keynes: theoretical methods and unfinished business', in D.Reese (ed.) *The Legacy of Keynes,* San Francisco CA: Harper and Row. Reprinted in Harcourt, G.C. (1992) *On Political Economists and Modern Political Economy,* ed. Claudio Sardoni, London: Routledge, 235 – 49.

Harcourt, G.C. and O'Shaughnessy, T. (1985) 'Keynes' unemployment equilibrium: some insights from Joan Robinson, Piero Sraffa and Richard Kahn', in G.C. Harcourt (ed.) (1985) *Keynes and his contemporaries,* New York: St Martin's Press, 3 – 41.

Hodgson, G. (1988) *Economics and Institutions,* Philadelphia PA, University of Philadelphia Press.

Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money,* London: Macmillan.

——(1937) 'The General Theory of Employment', *Quarterly Journal of Economics.* Reprinted in Keynes, J.M. (1973c) *The Collected Writings of John Maynard Keynes,* vol. XIV, London: Macmillan.

——(1971) *The Collected Writings of John Maynard Keynes,* vol. V, *A Treatise on Money,* London: Macmillan.

——(1973a) *The Collected Writings of John Maynard Keynes,* vol. VIII, *A Treatise on Probability,* London: Macmillan.

——(1973b) *The Collected Writings of John Maynard Keynes,* vol. XIII, London: Macmillan.

——(1973c) *The Collected Writings of John Maynard Keynes,* vol. XIV, London: Macmillan.

——(1979) *The Collected Writings of John Maynard Keynes,* vol. XXIX, London: Macmillan.

Kregel, J. (1980) 'Markets and institutions as features of a capitalistic production system', *Journal of Post-Keynesian Economics,* III, 1, fall: 32 – 48.

——(1981) 'On distinguishing between alternative methods of approach to the demand for output as a whole', *Australian Economic Papers,* vol. 20, no. 36, June: 63 – 71.

——(1987) 'Rational spirits and the post-Keynesian macrotheory of microfoundations', *De Economist,* 135, 4:520 – 32.

——(1990) 'Imagination, exchange and business enterprise in Smith and Shackle', in S.Frowen (ed.) (1990) *Unknowledge and Choice in Economics,* London: Macmillan, 81 – 95.

Lavoie, M. (1992) *Foundations of Post-Keynesian Economic Analysis,* Aldershot: Edward Elgar.

Lawson, T. (1985) 'Uncertainty and economic analysis', *Economic Journal,* 95, December: 909 – 27.

——(1991a) 'Keynes and the analysis of rational behaviour', in R.O'Donnell (ed.) (1991) *Keynes as Philosopher-Economist,* New York: St Martin's Press, 184 – 226.

——(1991b) 'Reply', in R.O'Donnell (ed.) (1991) *Keynes as Philosopher-Economist,* New York: St Martin's Press, 241 – 7.

——(1993) 'Keynes and conventions', *Review of Social Economy,* LI, 2, summer: 174 – 200.

——(1995) 'Economics and expectations', in S.Dow and J.Hillard (eds) (1995) *Keynes, Knowledge and Uncertainty,* Aldershot: Edward Elgar, 77 – 106.

Lecq, F. van der (1996) 'Conventions and institutions in coordination problems', *De Economist,* 144, 3:397 – 428.

Lewis, D. (1969) *Convention: A Philosophical Study,* Cambridge MA: Harvard University Press.

Littleboy, B. (1991) *On Interpreting Keynes,* London: Routledge.

Lo, K. (1996) 'Equilibrium in beliefs under uncertainty', *Journal of Economic Theory,* 71: 443 – 84.

McKenna, E. and Zannoni, D. (1993) 'Philosophical foundations of post-Keynesian economies', *Journal of Post-Keynesian Economics,* XV, 3, spring: 395 – 407.

Meeks, G. (1991) 'Keynes on the rationality of decision procedures under uncertainty: the investment decision', in G.Meeks (ed.) (1991) *Thoughtful Economic Man,* Cambridge: Cambridge University Press, 126 – 60.

Mirowski, P. (1986) 'Institutions as a solution concept in a game theory context', in Mirowski (ed.) (1986) *The Reconstruction of Economic Theory,* Boston MA: Kluwer, 241 – 63.

Neale, W. (1987) 'Institutions', *Journal of Economic Issues,* XXI, 3, September. Reprinted in M.Tool (ed.) (1988) *Evolutionary Economics, Volume I: Foundations of Institutional Thought,* Armonk NY: M.E. Sharpe, 227 – 56.

O'Donnell, R.M. (1989) *Keynes: Philosophy, Economics and Politics,* London: Macmillan.

Possas, M. (1987) *A Dinâmica da Economia Capitalista: Uma Abordagem Teórica,* Sa• o Paulo: Brasiliense.

——(1990) 'Rationality and regularity: towards an integration of micro and macro-dynamics', presented at the International Workshop in Post-Keynesian Economics, Knoxville TN, mimeo.

Rogers, C. (1993) *'The General Theory:* existence of a monetary long-period unemployment equilibrium', in G.C Harcourt and Riach: (eds) (1997) *A 'Second Edition of The General Theory,* vol. 1, London: Routledge: 324 – 42.

Rotheim, R. (1993) 'On the indeterminacy of Keynes' monetary theory of value', *Review of Political Economy,* 5, 2:197 – 216.

Runde, J. (1991) 'Keynesian uncertainty and the instability of beliefs', *Review of Political Economy,* 3, 2:125 – 45.

——(1994) 'The Keynesian probability-relation: in search of a substitute', in J.Davis (ed.) (1994) *The State of Interpretation of Keynes,* Boston MA: Kluwer, 245 – 51.

Scharfstein, D. and Stein, J. (1990) 'Herd behavior and investment', *American Economic Review,* 80, 3, June: 465 – 79.

Schotter, A. (1981) *The Economic Theory of Social Institutions,* Cambridge: Cambridge University Press.

Screpanti, E. (1995) 'Relative rationality, institutions and precautionary behaviour', in J.Groenewegen, C.Pitelis and S.Sjöstrand (eds) (1995) *On Economic Institutions,* Aldershot: Edward Elgar, 63 – 84.

Shackle, G. (1967) *The Years of High Theory,* Cambridge: Cambridge University Press.

——(1970) *Expectation, Enterprise and Profit,* London: Allen and Unwin.

——(1972) *Epistemics and Economics,* Cambridge: Cambridge University Press.

Simon, H. (1955) 'A behavioral model of rational choice', *Quarterly Journal of Economics,* 69:99 – 118.

——(1959) 'Theories of decision-making in economic and behavioral sciences', *American Economic Review,* 49:253 – 83.

Sugden, R. (1986) *The Economics of Rights, Cooperation and Welfare,* Oxford: Blackwell.

Townshend, H. (1937) 'Liquidity premium and the theory of value', *Economic Journal,* no. 47:157 – 69.

Vickers, D. (1979 – 80) 'Uncertainty, choice, and the marginal efficiencies', *Journal of Post-Keynesian Economics,* II, 2, winter: 240 – 54.

——(1994) *Economics and the Antagonism of Time,* Ann Arbor MI: University of Michigan Press.

Vromen, J. (1994) *Evolution and Efficiency,* Delft: Eburon.

Winslow, T. (1989) 'Organic interdependence, uncertainty and economic analysis', *Economic Journal,* 99, December: 1173 – 82.

——(1992) 'Keynes on rationality', in B.Gerrard (ed.) *The Economics of Rationality,* London and New York: Routledge, 91 – 122.

Young, H.P. (1996) 'The economics of convention', *Journal of Economic Perspectives,* 10, 2, spring: 105 – 22.

10
KEYNES ON RATIONALITY, EXPECTATIONS AND INVESTMENT

Michelle Baddeley

INTRODUCTION

In *The General Theory of Employment, Interest and Money* and in his 1937 *Quarterly Journal of Economics* article 'The general theory of employment', Keynes provides an alternative analysis of the links between past, present and future in investment decision-making to the complicated mathematical analyses of orthodox investment theory.[1] The essential difference between Keynes' analysis of investment and orthodox investment models lies not so much in the factors which are claimed as determinants of investment decisions but in different understandings of the nature of rational investor behaviour. The importance of a different understanding of rationality emerges because of the crucial role played by expectations in determining investment decisions: expectational assumptions are inevitably dependent upon an underlying conception of what rationality entails.

The differences between Keynes' conception of investor rationality and the orthodox treatment of rationality can be understood in terms of Simon's (1979) distinction between substantive and procedural rationality. Simon asserts that substantive rationality is associated with behaviour appropriate to 'the achievement of given goals within the limits imposed by given conditions and constraints' (Simon 1979:67). Substantive rationality assumptions underlie orthodox neoclassical models of expectations formation and investment decision-making. In these models, investors in fixed assets are described as self-interested profit-maximisers, forming precisely quantifiable expectations of future events. Whilst Jorgenson's (1963, 1971) early version of neoclassical theory abstracted from issues of rationality (expectations were assumed to be static), in refinements of Jorgenson's approach, expectations are of adaptive, extrapolative form and are assumed to be a function of past and current values of variables.

However, the poor empirical performance of early neoclassical models led to a marriage of the maximising assumptions of neoclassical theory with the rationality assumptions of the Rational Expectations Hypothesis (REH), for example in orthodox versions of q theory. In these models, expectations are

incorporated explicitly into the profit-maximisation problem which investors are assumed to face.

Dwyer, Williams, Battalio and Mason (1993) offer three differing interpretations of rational expectations:

1 Economic agents collect and use information efficiently.
2 Agents will behave as if the objective and subjective parameters that characterise the environment are the same. In other words, expectations are rational if the divergences between the expectations and realisations of a given variable are a white noise process.
3 Expectations, that is the subjective probability distribution of outcomes, given the same information set, tend to be distributed around the objective probability distribution of outcomes, i.e.:

$$_{t-1}X^{e}_{t} = E(\Omega_{t-1})$$

where $_{t-1}X^{e}_{t}$ is the expectation of variable X_t formed in period $t-1$ given the information available in period $t-1(\Omega_{t-1})$.

According to Muth (1961) the information set used by rational agents includes the relevant economic theory, and therefore modelling strategies which use economic theory to predict outcomes consistent with these versions of REH. Thus, with the REH, a concept of substantive, mechanical rationality similar to (although more sophisticated than) the rationality concepts of early neoclassical theory, is retained.

Simon (1979) comments that the substantive rationality approaches underlying these orthodox concepts 'freed economics from any dependence upon psychology'. However, the orthodox-style rationality assumption is of limited use in describing the complexity of real-world behaviour, particularly in describing decisions about fixed assets investment which tend to be non-repetitive. Simon argues that a procedural rationality assumption, in which decisions are 'the outcome of appropriate deliberation', is more appropriate to the analysis of real-world problems. This conception of rationality focuses on the reasoning process rather than the goals of a representative agent, and irrationality is defined as behaviour which is the outcome of impulse 'without an adequate intervention of thought' (Simon 1979:68).

Keynes' emphasis on the subjective determinants of investment, the limits to quantification and the role of conventional behaviour, fits broadly into a procedural description of rationality. In contrast to the orthodox analyses based on substantive rationality assumptions, Keynes argues that scientifictheories should be able to cope with real-world situations and should not force the facts to conform with theoretical assumptions (Keynes 1936:276). The facts of experience do not follow by necessity but are determined by the business environment and by psychological propensities. Therefore, at every stage of his

analysis, empirical observation, experience, generalisation and intuition, rather than logical necessity and precision, are his preferred analytical tools.

KEYNES' ANALYSIS OF RATIONALITY AND EXPECTATIONS

In Keynes' approach, psychological forces play a crucial role in his analysis, and Keynes argues that it is psychology of actual behaviour which vitiates classical and neoclassical analyses. 'Fundamental psychological factors': the propensity to consume, attitudes to liquidity and the schedule of the marginal efficiency of capital are, together with the wage unit and the quantity of money, the 'ultimate independent variables' in Keynes' system, determining the national income and quantity of employment (Keynes 1936:183). According to Keynes, when there is no basis for rational belief, behaviour is dictated by psychological motivations and non-rational forces, such as animal spirits and conventions. For example, mass psychology and conventions are of crucial importance in determining market valuations and interest rates (Keynes 1936:170).[2]

However, Keynes' ideas about rationality, convention and psychology and the extent to which they overlap are unclear. In both *A Treatise on Probability* and *The General Theory,* Keynes treats psychology as, in some sense, the contrary of rationality. However, Keynes does not distinguish adequately between mass psychology and conventional behaviour and at times he seems to treat them as distinct forces. It seems that conventions may be rational, non-rational or irrational/psychological depending on the nature of belief underlying them. Furthermore, conventions cannot be easily understood in terms of individual rationality because they reflect the overlap between individual and aggregate behaviour, where aggregate behaviour affects individual behaviour and vice-versa.[3] It is not rational for one individual to believe, in isolation, that the current situation can be projected into the future; but given that outcomes are determined by the aggregate behaviour, a conventional belief in the current situation as a guide to the future does have a reasonable basis. Therefore, there seems to be a distinction here between individual and social rationality.[4] To add to the confusion, Keynes often treats conventional behaviour as a rational or reasonable device in the face of uncertainty. Similarly, there seems to be some overlap between conventional and rational behaviour. An example is that he argues that the rate of interest is a highly conventional phenomenonrather than a highly psychological phenomenon (Keynes 1936:170, 203 – 4). Similarly, market valuations of investments are simultaneously the outcome of mass psychology, herd behaviour and/or convention.

This fuzziness in Keynes' ideas about rational versus conventional versus psychological forces has led to the development of divergent interpretations of his analysis. O'Donnell (1989, 1991), Littleboy (1990), Gerrard (1994), Meeks (1991), Bateman (1990), Skidelsky (1992) and Dow and Dow (1985) argue that

there is continuity in the development of Keynes' understanding of rational behaviour in his early work on probability and rational belief, as elucidated in Keynes 1921 and Keynes 1936. They argue that, in terms of Keynes' theory of rationality developed in Keynes 1921, the economic behaviour described in Keynes 1936 is rational, given lack of information about the future. In contrast, Shackle (1955, 1967, 1972), Winslow (1986) and Mini (1990) emphasise the subjectivism of economic behaviour and, to them, Keynes presents the volatile characteristics of the economy as the result of irrational forces, vicious tendencies, destructive fetishes and dangerous human proclivities. A further group of interpretations occupies the middle ground between the polar emphasis on rational calculation versus irrational animal spirits. Littleboy (1990), Runde (1997, 1991), Lawson (1981, 1985, 1995), Robinson (1979), Crotty (1992), Colander and Guthrie (1980), Minsky (1975), Davidson (1991a), Carabelli (1988) and Howitt (1997) argue that, in *The General Theory,* expectations and the conventions which determine them are sensible and reasonable rather than strictly rational or irrational.[5] Robinson (1979) views Keynes' conventions as non-rational and Littleboy (1990:34) argues that 'conventional behaviour lies between…two extremes, the fully rational and the fully irrational'. Conventions are not the same as customs/habits. They emerge from rational, purpose-oriented behaviour under uncertainty and promote coherent behaviour (Littleboy 1990: 271). According to Crotty (1992), social conventions 'create' and 'imagine' the missing data which links the logical chain connecting data to decisions.

The truth probably lies in the compromise implicit in the latter group of approaches. In *A Treatise on Probability,* Keynes does not preclude irrational belief and action and the mixture of rational, non-rational and irrational forces which determine economic behaviour in *The General Theory* can be understood in terms of his ideas in *A Treatise on Probability*. An understanding of the development of Keynes' ideas about rationality, expectations and uncertainty allows the resolution of some of the seeming contradictions and inconsistencies in Keynes' analysis of investment.

Rationality in *A Treatise on Probability*

In *A Treatise on Probability,* Keynes argues that probabilistic knowledge forms the basis of rational belief and arises from direct knowledge of aprobability-relation. The probability-relation is the rationally based connection between a body of evidence and the primary proposition. Direct acquaintance with the probability-relation gives rise to direct knowledge of a secondary proposition, which describes the rationally justifiable degree of belief in a conclusion, or 'primary proposition' (Keynes 1921:10 – 12). Whilst what is probable is what is rational to believe, it is not rational to believe the probable to be true, only to believe it in preference to other less probable beliefs (Keynes 1921:339). Keynes views probability as the 'guide of life'. He therefore argues that there is a

fundamental connection between rational belief and rational action: the hypothesis which is probable is the hypothesis on which it is rational to act (Keynes 1921:339 – 40).

Keynes' theory of rationality, like his theory of investment, is a mixture of subjective and objective elements. Although in one sense the probability-relation is objective and any person operating from the same knowledge base will arrive at the same secondary proposition, in another sense it is subjective because knowledge/evidence varies across individuals and over time, and the rationally justifiable degrees of belief in a given conclusion will change accordingly. Keynes emphasises this relational aspect to probability. All probability must be defined in relation to the evidence which underlies it; to say merely that some proposition is probable is 'elliptical' (Keynes 1921: 8). Keynes' insight about this relational aspect of probability is important for the practical, common-sense understanding of probability and the formation of expectations. Probabilistic knowledge cannot be defined in terms of absolutes. Just as our estimates of probability are relative to our knowledge of other propositions, they are also relative to our capacity for logical insight (Keynes 1921:18 – 19). Keynes is critical of statistical, frequency theories of probability which argue that there is an objective probability distribution of outcomes, revealed via experience (Keynes 1921: 94). Frequency theories form the basis of the orthodox REH in that agents' subjective probability distributions are assumed to coincide with objective probability distributions.

Ramsey observes that Keynes, by arguing that the ability to discern probability relations is constrained by the limits on human reason, undermines the objective basis of his theory. If probability relations are to be objective they must be 'independent of the human mind' (Ramsey 1931:164). Braithwaite (1973) also argues that Keynes' attempt at an axiomatic theory of probability has formal defects. According to Braithwaite difficulties emerge in the assumption that some probabilities are not measurable. Also the relativity of probability relations to human reason 'throws over entirely [Keynes'] doctrine of specific objective probability-relations' (Braithwaite 1972:xxi). O'Donnell defends Keynes' ideas from these criticisms and observes that Keynes' emphasis on limited human insight is consistent with his overall theory of probability (O'Donnell 1989:64 – 6).[6]

The criticisms of Ramsey and Braithwaite have their roots in the looseness of Keynes' expression. The incompleteness of Keynes' arguments in *A Treatise on Probability* hinders the interpretation of his arguments, but it seems that he was arguing that logical ability and the logical rules which enable the perception of a probability relation form part of the premises which are a given individual's knowledge. All agents therefore have equal capacity for rationality entailed by the logical rules with which they are familiar. An analogy can be drawn with mathematics. If you do not know the rules of addition it is not necessarily rational to believe that 2 plus 2 equals 4. If you do know the rules of addition it is

irrational to believe otherwise. This does not prevent some individuals being less consistent in the application of these rules and therefore more likely to fall into the trap of irrational belief. There may be no rationally justifiable probability relation because our knowledge base does not include the necessary logical rules and information which would allow the probability relation to be discovered. In such circumstances action must be driven by motivations other than rational belief.

Keynes' understanding of irrational belief involves the failure of individuals in a given situation to exercise their logical tools and use the information available to them. Irrational belief develops when people believe in a non-existent probability relation or fail to perceive a probability relation even though, with more careful thought, they should be able to perceive one (Keynes 1921:34 – 5). An example of irrational behaviour based on irrational belief would be relying on psychological motivations rather than reasoning power, even when distinct probability relations can be formed.

The quantification of probability

Another crucial aspect of Keynes' theory of rationality is the extent to which rational belief depends on the precise quantification of probabilities. Throughout *The General Theory,* Keynes emphasises the difficulties in forming precise mathematical expectations of future events. He argues that the problem with Classical and neoclassical analyses is that they assume that relevant facts about the future are more or less known, and that expectations can be precisely quantified. Keynes argues that orthodox analysis fails to distinguish calculable risk from inherently unquantifiable uncertainty. In orthodox theory, risks are assumed to be capable of 'exact actuarial computation' and exist without unquantifiable uncertainty. Underlying this approach is the implicit adoption of a calculus of probability, in which not only risk but also uncertainty can be precisely quantified. The principles of behaviour adopted because of a need for action are misinterpreted as having a precisely quantifiable, rational basis, and therefore factors such as 'utter doubt, precariousness, hope and fear' are underestimated by orthodox theory (Keynes 1937:224).[7]

In *A Treatise on Probability,* Keynes' distinction between risk anduncertainty can be understood in terms of the different limits in different circumstances on the quantification of probability relations. Thus Keynes develops a general theory of rationality which can encompass *all* types of probabilistic knowledge, whether quantifiable or not. Some probability relations will be reducible to a single number, others will only be ordinally comparable and yet others will not be quantifiable at all. Keynes argues that this last category of probability relations has mistakenly been understood in terms of unknown probabilities when the impossibility of quantifying probability relations has not been recognised. In the case of unquantifiable probabilities we can only say that a

conclusion based on the evidence is not impossible, not certain, but nonetheless cannot be compared with other probabilities. Thus Keynes argues that theories of rationality which emphasise precisely quantifiable probability are a special case within his general theory of probability. Keynes also argues that the numerical quantification of probability is possible only in a small minority of cases.

The 'principle of indifference', first formulated by Bernoulli, is essential to the numerical measurement of probability. Keynes states that 'the principle of indifference asserts [that] equal probabilities must be assigned to each of several alternatives, if there is an absence of positive ground for assigning unequal ones' (Keynes 1921:45). Therefore, the probabilities of mutually exclusive, exhaustive alternatives will sum to unity, and the probability of each of two hypotheses, given the application of the principle of indifference, can be assumed to equal 1/2. Keynes explores numerous examples where quantification of probabilities has taken place via a misguided application of this principle of indifference, and uses these examples to elaborate upon and refine the principle (Keynes 1921:44, 69).

The existence of probability relations which are not precisely quantifiable suggests that different people may act in different ways even if they operate from the same knowledge base, because subjective judgements will play a role in the formation of beliefs. The insurers in the *Treatise on Probability,* for example, will derive premiums which allow margins for error and therefore, whilst insurance premiums appear to be based on a precise quantification of the probability of events, they in fact have an arbitrary component (Keynes 1921:23).

In 1932 Keynes seemed to maintain this understanding of the role of quantification within margins, arguing that 'a definition can often be vague within fairly wide limits and capable of several interpretations differing slightly from one another and still be perfectly serviceable and free from serious risk of error' (Keynes 1979:36). Similarly in *The General Theory,* changes in news will be interpreted differently by different individuals. The bulls and bears dealing in bond markets each hold a different set of beliefs about the likelihood of a change in interest rates, which reflects their subjective judgements.

Another example of the different degrees of quantification arises in thecontext of supplementary costs and windfall losses. Keynes does discuss insurable risk as that part of supplementary costs which are 'sufficiently regular to be commonly regarded as 'insurable risks' (Keynes 1936:56). But, as discussed above, insurable does not necessarily imply precisely quantifiable: 'no unique principle can be established for the estimation of supplementary cost'. However, the expected value of supplementary cost is assigned a definite quantity, although it may be re-estimated over the life of equipment, as expectations change (Keynes 1936:57 – 8). In contrast, windfall losses are involuntary, 'unforeseen changes in market values, exceptional obsolescence or destruction by catastrophe' and are subject to fundamental uncertainty rather than quantifiable risk. Other examples include: the quantification of user cost: it is 'at least equal to…and it may be more' than the present value of the opportunity to

postpone replacement (Keynes 1936:70) and risk margins between the rate of interest and expectation of yield required by entrepreneurs before they will borrow (Keynes 1936:144 – 5).

Weight and the state of confidence

Weight plays a crucial role in Keynes' theory of rationality, even though he appears confused and unclear about the concept. Keynes argues that, whereas the probability relation reflects a balance of favourable and unfavourable evidence, weight is the absolute amount of relevant evidence (Keynes 1921:77). A probability relation with greater weight therefore rests on a more substantial basis. Weight cannot be explained in terms of probability because an argument with greater weight is not more likely to be right: all probability relations are equally justifiable given the evidence upon which they are based. Weight and probability are therefore independent properties.

Keynes argues, along with Bernoulli, that we should be guided by the probability with the greatest evidential weight. However, the implications of this maxim are unclear and there is no clear principle about how far weight should be strengthened. In some cases rational action should involve discovering probability relations with greater weight by acquiring more information; in others, new information is not worth the trouble of finding it. In this, Keynes' discussion of weight seems to parallel the cost-benefit approach to information searches (Keynes 1921:82 – 5). Although Keynes is unsure about the practical significance of weight, he does argue that the rationally preferable course of action is that based on probability relations with the least 'risk' as well as the most weight. However, Keynes argues that it is difficult to assess the relative influences of weight and risk and suggests a 'conventional coefficient' of weight and risk (Keynes 1921:347 – 8). His derivation of this coefficient seems anomalous given his emphasis on the limits to quantifying probability and weight.

The concept of weight resurfaces in *The General Theory,* in Keynes' discussion of the state of confidence. Keynes associates liquidity premiawith judgements of weight and risk premia with judgements of probability (Runde 1997:6). Therefore it seems that a lower weight, associated with a higher liquidity premium, reflects a weakening state of confidence. The state of confidence plays a crucial role in Keynes' theory of investment because the rate of investment is affected by the state of confidence indirectly via its effects on the marginal efficiency of capital. Keynes points out that the state of confidence is not analysed in much detail in economic analysis but that our understanding of it must be based on empirical observation rather than *a priori* theory. Gerrard argues that '[Keynes'] previous insight on the practical importance of weight is given specific form as the state of confidence' (Gerrard 1994:334). Minsky also equates weight with confidence (1975:65). However, the state of confidence

appears to be distinct from weight. It is a diffuse, global force, whereas a judgement of weight will apply only to a given probability relation. It seems that the connection between the two concepts arises because the state of confidence determines whether or not judgements of weight can be made. When the state of confidence is buoyant, judgements of weight are possible. When the state of confidence becomes more fragile, judgements of weight are more difficult. In the latter case, there will be no guide to rational action and non-rational forces will predominate.

Vague knowledge

In *The General Theory,* the beliefs determining long-term expectations of the yields of investments arise out of uncertainty about the future of investments. Vague knowledge or uncompleted knowledge, associated with indistinct probability relations, lies at the core of Keynes' concept of uncertainty about the future.[8] Uncertainty does not arise because probability relations cannot be quantified, but because they are not distinctly perceived or because no probability relation exists: there is not enough information to form rational belief. As Crotty succinctly states it: there is an 'empty space' in the logical chain connecting data to decisions (Crotty 1992:486). Thus vague knowledge, low weight and uncertainty seem to be closely connected, and the interconnections between uncertainty and vague knowledge can be understood in terms of ideas from *A Treatise on Probability*.

Whilst, in *A Treatise on Probability* (1921:17 – 18) Keynes does not appear to have a clear understanding of the nature of vague knowledge, his ideas about vague knowledge resurface in *The General Theory* in his treatment of behaviour under uncertainty. There is considerable evidence that Keynes still understood vague knowledge in terms of ideas from *A Treatise on Probability,* especially weight.[9]

However, a more complete interpretation of Keynes' ideas would be that, by the time of writing *The General Theory,* Keynes understood the word 'vague' in at least two senses: not only vague, imprecise concepts, but also the vague knowledge available to an individual when forming belief.Examples of vague concepts in *The General Theory* include user cost (1936: 71), expectations of a distant future (148), liquidity (240), the quantity of output and capital equipment and the general level of prices (39, 43).

Vague knowledge of vague concepts may also arise not only out of a concept's inherent vagueness but also because of lack of information about the future. Vague knowledge may also apply to concepts which we can conceive of in precise terms, e.g. wages and labour, because lack of information prevents us forming precise expectations of what will happen to them. Keynes argues that in the real world we will only have vague knowledge about most of the consequences of our actions. The ubiquity of vague knowledge about crucial

determinants of investment decisions and expectations does not prevent practical decision-making but again means that analyses of decision-making based on logic and precision will be inappropriate.

Vague knowledge reflects insufficient information and/or inherent limitations to logical power, and therefore is associated with probability relations with a weight so low that the probability relation cannot be distinctly perceived. The role of vague knowledge within Keynes' theory of rationality is unclear. Given no distinct probability relation it is difficult to say whether beliefs thereupon formed are irrational or rational. In a situation of vague knowledge, probability relations are indistinct and Keynes argues that one of the roles of probabilistic knowledge is to convert such vague relations into more distinct ones (possibly via judgements of weight, although Keynes does not state this explicitly). A theory of probability can illuminate the characteristics of rational belief under conditions of uncertainty and vague knowledge because 'logical systems of probability enable us to know the relations, which cannot be easily perceived, by means of other relations which we can recognise more distinctly—to convert, in fact, vague knowledge into more distinct knowledge' (Keynes 1921:57).

On the other hand, Keynes appears to make a connection between vague knowledge, unknown probabilities and irrational belief:

> To say, then, that a probability is unknown ought to mean that it is unknown to us through a lack of skill in arguing from given evidence. The evidence justifies a certain degree of knowledge, but the weakness of our reasoning power prevents us knowing what this degree is.
>
> (Keynes 1921:34)

Indistinct probability relations and vague knowledge therefore lie in a twilight zones between no probability relation and unknown probability relations and between rational and irrational belief. What is clear is that actions in situations characterised by vague knowledge, where probabilistic knowledge does not allow more distinct knowledge to be derived, may beeither rational, non-rational but reasonable, or irrational.[10] Under these conditions, if people act from other motivations, rather than a firm (but necessarily misplaced) belief, then action may be non-rational but nonetheless reasonable, given no basis for rational belief. To act from belief when there is no basis for rational belief must necessarily be irrational. The only rational *action,* given vague knowledge, is to rely on other motivations. These understandings of rationality, irrationality and non-rationality can illuminate our understanding of the rationality of investor behaviour based on conventions.

KEYNES ON RATIONALITY, EXPECTATIONS AND INVESTMENT

The elements of *A Treatise on Probability* discussed above may be reconciled with elements of *The General Theory* to give a unique analysis of investment which is based on Keynes' understanding of the role of (procedural) rationality in expectations formation. The role of rational thought processes in decision-making will vary according to the circumstances which a given investor faces. At times, when it is possible to do so, rational devices are used; at other times other forces take over. Investor behaviour is an amalgam of different decision-making approaches and cannot be strictly classified. Rational, irrational and non-rational behaviours all enter the investment decision-making processes, but at different stages.

Keynes focuses on a number of different influences determining investment decisions. These influences can be separated into objective factors (e.g. the marginal efficiency of capital and interest rates) and subjective factors (e.g. expectations and animal spirits).[11] Keynes' ideas about rationality are most crucial to understanding the latter category and, in particular, to understanding the way in which expectations about investments' likely future performance are formed. However, Keynes' distinction between short-term and long-term expectations further complicates his analysis. Keynes does not analyse the formation of short-term expectations in the context of vague knowledge and uncertainty. The formation of short-term expectations is based on rational belief, and in this sense is more compatible with an assumption of rational expectations (Howitt 1997). Entrepreneurs, when making decisions about the scale of their production, will entertain not one expectation but a bundle of expectations held with varying degrees of probability and weight. However, the entrepreneur will act as if his or her behaviour is the outcome of one undoubting expectation held with certainty, operating on the expectation with the greatest weight (*ibid.:* 24). Thus rational judgements of weight underlie the convention of assuming that current events and most recent realised results can proxy short-term expectations (Keynes 1936:51). The weight of current events in forming short-termexpectations outweighs the costs of working out more complicated expectations *'de novo'* and the convention of using most recently realised results as a substitute for expectations emerges.

In contrast, the formation of long-term expectations is based upon the convention of assuming that the existing situation will persist (Keynes 1936: 162 – 3). This convention appears to combine rational and non-rational elements, and Keynes argues that it is rational because:

> We are assuming, in effect, that the existing market valuation…is uniquely correct in relation to our existing knowledge of the facts which will influence the yield of an investment, and that it will only change in proportion to changes in this knowledge.

(*ibid.:* 152)

Judgements of weight may also take place when

> it is reasonable to be guided to a considerable degree by the facts about which we feel somewhat confident, even though they may be less decisively relevant to the issue than other facts about which our knowledge is vague and scanty.
>
> (*ibid.:* 148)

However, this convention does not seem to be soundly based in rational belief. It is maintained even though experience has shown that it is unlikely that the 'existing state of affairs will continue indefinitely' (*ibid.:* 152).

Psychological factors also play a role, and Keynes argues that the 'conventional valuation which is established as the outcome of the mass psychology of a large number of ignorant individuals is liable to change violently' (*ibid.:* 154). Lawson (1995) argues that the conventions in Keynes' analyses have both a reasonable and a psychological element: they become self-fulfilling prophecies and therefore to assume that they will continue becomes the most reasonable thing to believe, once the convention is established. Also, for an ignorant individual it is reasonable to rely on conventions, because other economic actors may be acting on better information and 'we endeavour to fall back on the judgement of the rest of the world which is perhaps better informed' (Keynes 1936:217). However, the convention is also maintained because of psychological factors: people prefer stable routines, and conventions lull the anxiety created by uncertainty about the future (Lawson 1995; Earl 1983).

Keynes (1979:294) argues that in making decisions a large number of alternatives exist and none are obviously more rational than others. If it is not possible to rank these alternatives via judgements of weight, then we fall back on other motivations which are not rational in the sense of beingevaluated in terms of consequences, but instead are determined by instinct, preference and desire.

Agents are nonetheless doing the best that they can in the circumstances. In the context of investor behaviour, if the ubiquity of vague knowledge prevents the formation of rationally based estimates of prospective yield, then investors must rely upon non-rational forces in forming their estimates of prospective yields. Behaviour becomes the outcome of animal spirits and spontaneous optimism/pessimism. Relying on such motivations will not be irrational given that there is no basis for rational belief. This conception of rational behaviour is consistent with Simon's concept of procedural rationality in which agents' actions are the outcome of appropriate deliberation (Simon 1979).

Whilst many elements underlying investment decisions may have a rational or reasonable basis, irrationality may also play a role in investment decision-making. As discussed above, Keynes seems to argue that irrational belief is

distinct from non-rational belief because it is based on a misjudged or unknown probability relation, for example when assuming that a probability relation exists when none exists. In terms of *A Treatise on Probability,* to follow a convention whilst recognising that no objective basis for belief exists is not necessarily irrational, but to follow a convention because one falsely believes it has an objective basis is irrational because it reflects incorrect perceptions of probability relations. An example arises in the determination of the rate of interest, which Keynes describes as a 'highly conventional phenomenon' (Keynes 1936:203). However, he also argues that the rate of interest 'established by convention is thought to be rooted in objective grounds much stronger than convention' (Keynes 1936:204). In other words, although the rate of interest is established by convention, people nonetheless perceive it to have an objective basis and therefore, in terms of the language of *A Treatise on Probability,* they think that they perceive a probability relation which in reality does not exist. The belief that the interest rate has its roots in objective grounds may lead to an inappropriate complex of interest rates prevailing due to these mistaken beliefs. This may lead to sub-optimal levels of employment being maintained over periods of time as the conventional, relatively stable but high long-term rate of interest prevails (Keynes 1936:203 – 4).

Keynes also allows irrational action in terms of ignoring rationally formed belief. An example relates to the behaviour of entrepreneurs. Keynes argues that entrepreneurs fall in with the ideas of the market even though they may themselves be better instructed about the prospects of their individual investments. They seem to have formed some sort of rationally based belief about prospective yields. Rational entrepreneurs, who are concerned with prospective yields rather than forecasting the market, would act on this belief. To ignore it is irrational unless entrepreneurs are also concernedabout the loss of liquidity which is the cost of diverging from average opinion.

Other mistakes made by agents are not necessarily irrational. They may be based on an objective probability relation and just turn out ex post to be misplaced, for example the fluctuations in inventories arising from mistaken short-term expectations (Keynes 1936:332) and estimates of user cost (*ibid:* 290). Similarly, entrepreneurs may assess, on the basis of hindsight, that they have made mistakes, but these mistakes are not irrational in the sense of *A Treatise on Probability* if there was no basis for rational belief to start with.

CONCLUSION

Keynes presents a complex analysis of the nature of the inducement to invest, and an equally complex analysis of the nature of economic behaviour under conditions of uncertainty. Both analyses contain an amalgam of objective and subjective factors, and are characterised by inconsistencies and omissions which sometimes makes the interpretation of his arguments difficult and to a certain

extent arbitrary. His contribution to the understanding of the psychology of investor behaviour is nonetheless crucial. Overall, investor behaviour is described by Keynes as a mixture of rational, irrational and non-rational elements, with the individual influence of the elements dependent upon whether judgements of probability and weight can convert vague knowledge into more objective belief. If beliefs are irrationally based, or if there is no rational basis for belief, action may be the outcome of purely psychological motivations. This analysis of behaviour fits well with an intuitive understanding of real-world decision-making.

NOTES

1 See Chirinko (1993) and Baddeley (1995) for surveys of orthodox investment modelling strategies.
2 Harcourt and Sardoni (1996) warn, however, that to lay too much emphasis on individual psychology may be misplaced because, in general, Keynes addresses issues of social psychology and conventional behaviour rather than individual psychological responses.
3 The overlap between individual and aggregate behaviour and the problem of precise quantification have led to a belief that Keynes' investment demand schedule is an inappropriate construction which is in many senses unnecessary for, and contradictory with, the rest of Keynes' theory. See also Davis (1996) on the interdependence of individual economic agents in *The General Theory*.
4 There is also a distinction between individual and institutional rationality. Hodgson (1985) argues that Keynes' emphasis on the behaviour of the entrepreneur as a determinant of investment decisions is misplaced because the agency forming expectations within the firm is rarely a single commanding individual (1985:17). Firms' decisions reflect expectations and objectives of the largenumber of people. These ideas have been developed in the behavioural theory of the firm (see Cyert and March 1963; Simon 1979; Morris 1974).
5 Skidelsky (1992:86 – 9) presents a survey of the different interpretations of Keynes' analysis of economic behaviour, which centres around the extent to which these interpretations focus on rational, irrational or non-rational forces.
6 O'Donnell argues that Keynes believed in a superior capacity for rationality of those with greater logical ability (O'Donnell 1989:66). In O'Donnell's interpretation, capacity for rationality is dependent on the subjective element of logical capacity and thus any sense of an objectively based probability relation is lost. Harcourt and Sardoni (1996:7) argue that O'Donnell's account of Keynes' philosophy and economics, although masterly, suffers in its attempts to be 'too systematic, explicit and formally logical and leaves too little room for Keynes' leaps of intuition'.
7 Davidson (1991b, 1995) argues that orthodox theory, in contrast to the theories of Shackle and Keynes, treats economic agents as robotic decision-makers reacting to events and thus fails to explain the creative role that entrepreneurs play in determining the economic future

8 Uncompleted knowledge is the knowledge 'such as we have in ordinary thought by passing from one proposition to another without being able to say what logical relations, if any, we have perceived between them' (Keynes 1921:14). Thus the term seems to be used be Keynes in much the same sense as vague knowledge. O'Donnell (1989) argues that under uncertainty, knowledge is vague and probability relations are indistinctly perceived. O'Donnell therefore seems to associate vague knowledge with unknown probabilities. Although there are elements in Keynes 1921 which justify this view, overall Keynes seems to imply that unknown probabilities are unknown because of a failure to exercise reasoning power: an objective probability relation does exist but the irrational agent has failed to perceive it (1921:34 – 5).

9 In a footnote (Keynes 1936:148) he directly refers the reader to *A Treatise on Probability* and in his 1937 article when he states that: 'our knowledge of the future is fluctuating, vague and uncertain' (1937:216). However, Coates (1997) argues that Keynes' conception of the meaning of 'vague' underwent a transformation between the writing of *A Treatise on Probability* and *The General Theory,* reflecting the influence of the Wittgensteinian revolution in philosophy. Coates argues that by the time he wrote *The General Theory,* Keynes had come to associate vagueness with lack of precision rather than indistinct probability relations. See also Davis (1997) on Wittgenstein's influence on Keynes' understanding of conventions.

10 Rational action has been confused with rational belief. There is a distinction, and Keynes warns that 'we may sometimes confuse the practical certainty attaching to the class of beliefs upon which it is rational *to act* [emphasis added] with the utmost confidence, with the more wholly objective certainty of logic' (Keynes 1921:275).

11 This separation of objective and subjective factors abstracts from the interplay between subjective factors, such as expectations, and objective factors, such as the marginal efficiency of capital.

REFERENCES

Baddeley, M.C. (1995), 'Rationality, expectations and investment: the theory of Keynes vs neoclassical theory', Ph.D. thesis, University of Cambridge.

Bateman, B.W. (1990), 'The elusive logical relation: An essay on change and continuity in Keynes's Thought', in D.E.Moggridge (ed), *Perspectives on the History of Economic Thought,* vol. IV, Aldershot: Edward Elgar.

Braithwaite, R.B. (1973) 'Editorial foreword', in *The Collected Writings of John Maynard Keynes,* vol. VIII, London: Macmillan, xv – xxii.

Bulmer, M.G. (1989) *Principles of Statistics,* New York: Dover.

Carabelli, A. (1988) *On Keynes' Method,* London: Macmillan.

Chirinko, R.S. (1993) 'Business fixed investment spending: modelling strategies, empirical results and policy implications', *Journal of Economic Literature,* XXXI: 1875 – 1911.

Coates, J. (1990) 'Ordinary language economics: Keynes and Cambridge philosophers', Ph.D. thesis, University of Cambridge.

——(1997) 'Keynes vague concepts and fuzzy logic', in G.C.Harcourt and P.Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 2, London and New York: Routledge, 244 – 57.

Colander, D.C. and Guthrie, R.S. (1980) 'Great expectations: what the Dickens do "Rational expectations" mean?', *Journal of Post-Keynesian Economics,* 3, 2: 219 – 34.

Crotty, J.R. (1992) 'Neoclassical and Keynesian approaches to the theory of investment', *Journal of Post-Keynesian Economics,* 14, 4:483 – 96.

Cyert, R.M. and March, J.G. (1963) *A Behavioural Theory of the Firm,* Englewood Cliffs NJ: Prentice-Hall.

Davidson, P. (1991a) 'Comment', in R.M.O'Donnell (ed.) *Keynes as Philosopher-Economist,* proceedings of the ninth Keynes seminar, University of Kent at Canterbury, London: Macmillan: 61 – 72.

——(1991b) *Inflation, Open Economies and Resources,* London: Macmillan.

——(1995) 'Uncertainty in economies', in S.Dow and J.Hillard (eds) (1995) *Keynes Knowledge and Uncertainty,* Aldershot: Edward Elgar, 107 – 16.

Davis, J.B. (1997) 'J.M.Keynes on history and convention', in G.C.Harcourt and P. Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 2, London and New York: Routledge, 203 – 21.

Dow, A. and Dow, S. (1985) 'Animal spirits and rationality', in T.Lawson and M.H. Pesaran (eds) (1985) *Keynes' Economics: Methodological Issues,* Beckenham: Croom Helm.

Dwyer, G.P., Williams, A.W, Battalio, R.C. and Mason, T.L. (1993) 'Test of rational expectations in a stark setting', *Economic Journal,* 103:586 – 601.

Earl, P.E. (1983) *The Economic Imagination: Towards a Behavioural Analysis of Choice,* New York: Sharpe.

Gerrard, B. (1994) 'Beyond rational expectations: a constructive interpretation of Keynes' analysis of behaviour under uncertainty', *The Economic Journal,* 104: 327 – 37.

Harcourt, G.C. and Sardoni, C. (1996) 'The general theory of employment, interest and money: three views', in P.Arestis (ed.) *Keynes, Money and the Open Economy: Essays in honour of Paul Davidson,* vol. 1, Aldershot: Edward Elgar, 1 – 13.

Hodgson, G.M. (1985) 'Persuasion, expectations and the limits to Keynes', in T. Lawson and M.H.Pesaran (eds) (1985) *Keynes' Economics: Methodological Issues,* Beckenham: Croom Helm.

Howitt, P. (1997) 'Expectations and uncertainty in contemporary Keynesian models', in G.C.Harcourt and P.Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 1, London and New York: Routledge, 238 – 60.

Jorgenson, D.W. (1963) 'Capital theory and investment behaviour', *American Economic Review,* 53:247 – 59.

——(1971) 'Econometric studies of investment behaviour: a survey', *Journal of Economic Literature,* 9 (4):1111 – 47.

Keynes, J.M. (1921) *A Treatise on Probability,* London: Macmillan.

——(1936) *The General Theory of Employment, Interest and Money,* London: Macmillan.

——(1937) 'The general theory of employment', *Quarterly Journal of Economics,* 51: 209 – 23.

——(1979) *The General Theory and After: A Supplement. Collected Writings of John Maynard Keynes,* vol. XXIX, London: Macmillan.

Lawson, T. (1981) 'Keynesian model building and the rational expectations critique', *Cambridge Journal of Economics,* 5:311 – 26.

——(1985) 'Uncertainty and economic analysis', *Economic Journal,* 95:902 – 27.

——(1995) 'Economics and expectations', in S.C.Dow and J.Hillard (eds), *Keynes, Knowledge and Uncertainty,* Aldershot: Edward Elgar, 77 – 106.

Littleboy, B. (1990) *On Interpreting Keynes: A Study in Reconciliation,* London and New York: Routledge.

Meeks, J.G. (1991) 'Keynes on the rationality of decision procedures under uncertainty: the investment decision', in J.G.Meeks, (ed.) *Thoughtful Economic Man,* Cambridge: Cambridge University Press.

Mini, P. (1990) *Keynes, Bloomsbury and The General Theory,* London: Macmillan.

Minsky, H.P. (1975) *John Maynard Keynes,* New York: Columbia University Press.

Morris. D.J. (1974) 'The structure of investment decisions', D. Phil, Oxford University.

Muth, J. (1961) 'Rational expectations and the theory of price movements', *Econometrica,* 29:315 – 35.

O'Donnell, R.M. (1989) *Keynes: Philosophy, Economics and Politics,* London: Macmillan.

——(1991) 'Keynes on probability, expectations and uncertainty', in R.M.O'Donnell. (ed.) *Keynes as Philosopher-Economist,* proceedings of the ninth Keynes seminar, University of Kent at Canterbury, London: Macmillan, 3 – 60.

Pasinetti, L.L. (1997) 'The Marginal efficiency of investment', in G.C.Harcourt and P.Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 1, London and New York: Routledge, 198 – 218.

Ramsey, F. (1931) 'Truth and probability', in F.Ramsey, *The Foundations of Mathematics and Other Logical Essays,* edited by R.B.Braithwaite, London: Kegan Paul, Trench, Trübner and Co., 156 – 98.

Robinson, J. (1979) 'What has become of the Keynesian revolution?', in Robinson, J. *Collected Economic Papers,* vol. 5, Oxford: Blackwell.

Runde, J.H. (1991) 'Keynesian uncertainty and the instability of beliefs', *Review of Political Economy,* 3, 2:125 – 45

——(1997) 'Keynesian methodology', in G.C.Harcourt and P.Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 2, London and New York: Routledge, 222 – 43.

Shackle, G.L.S. (1955) *Uncertainty in Economics and Other Reflections,* Cambridge: Cambridge University Press.

——(1967) *The Years of High Theory,* Cambridge: Cambridge University Press.

——(1972) *Epistemics and Economics,* Cambridge: Cambridge University Press.

Simon, H.A. (1979) 'From substantive to procedural rationality', in F.H.Hahn and M.Hollis (eds) (1979) *Philosophy and Economic Theory,* Oxford: Oxford University Press, 65 – 86.

Skidelsky, R. (1992) *John Maynard Keynes: The Economist As Saviour, 1920 – 37,* London: Macmillan.

Winslow, E.G. (1986) 'Keynes and Freud: psychoanalysis and Keynes' account of the "animal spirits" of capitalism', *Social Research,* 53, 4:549 – 78.

11
KEYNES' COMMON-SENSE ECONOMICS
A criticism of Coates' argument

F.V.Comim

Interest in the philosophical foundations of Keynes' thought has been developed along two distinct lines. Initially, we saw the emergence of broad accounts of the philosophical foundations of Keynes' thought built upon his 1921 *A Treatise on Probability*. We should mention here the pioneer works of Meeks, Carabelli, O'Donnell and Lawson scrutinising the influence of Cambridge analytical philosophers, in particular that of G.E.Moore, on Keynes' earlier intuitionist ethics. Subsequently, investigations such as those of Coates, Cottrell, Skidelsky and Davis, have stressed Keynes' acceptance of Wittgenstein's later philosophy and challenged the assertion that Keynes' early philosophical ideas would have slipped without any friction into his economic thinking. It is within this context that we shall frame and assess the argument developed in John Coates' book *The Claims of Common Sense* (1996), in particular the claim that Keynes would have formulated an 'ordinary language economies'. We argue that, although this book is relevant in that it addresses an important issue and inspires a constructive critical reaction, the cogency of the basic arguments it proposes might be called into question. Ordinary language philosophy is not, as proposed by Coates, the central point of Keynes' economics. The foundation of Keynes' economics is his own common-sense philosophy resulting from his belonging to the Marshallian tradition and practising as an economist, influenced but not compelled by external philosophical conundrums. Coates' argument is certainly an important contribution to the literature emphasising Wittgenstein's influence on Keynes' philosophical development. However, Coates' analysis could be extended to a position more illuminating and useful for the methodology of economics if, by reorganising the theoretical elements he provides, we could set out a defence of Keynes' common-sense economics.

THE GENERAL ARGUMENT

The general problem addressed by Coates is the gap existing today, within the dominant contemporary philosophical paradigms, between the theory and practice of social sciences. The solution he proposes lies between the formal semantics of analytic philosophy and the textual nihilism of post-structuralism, in

'the middle route' of common sense, vague concepts and ordinary language of social scientists. In order to provide the grounds for the proposed solution, Coates attempts to track its roots to the Cambridge philosophy between the wars, with special emphasis on G.E.Moore's defence of common sense, F.Ramsey's and L.Wittgenstein's analysis of vagueness, and J.M.Keynes' application of these ideas to economics.

The book contains six chapters and a conclusion. In the first chapter Coates surveys the history of the notion of common-sense philosophy and relates it to the logic of vagueness, which according to him 'emerged as an integral part' of common-sense analysis. More specifically, he argues that ordinary language and common sense are closely tied and that in the light of contemporaneous philosophy, under a Quinian perspective, common sense is tantamount to 'the counsel of conservatism in the choice of hypotheses'. In the second chapter, Coates analyses the conflict in Cambridge philosophy between the early analytic tradition and the later Wittgenstein's analysis of vagueness and ordinary forms of discourse. In the third chapter, he describes the influence of Moore's ethics on Keynes' philosophical and methodological writings and associates it with some aspects of the method of *The General Theory*. In the fourth chapter, Coates presents Wittgenstein's criticisms of Moore's phenomenological approach and claims that the cogency of these criticisms led Keynes to accept key Wittgensteinian notions, in particular the notion of vagueness, which became central to Keynes' economics. The main argument Coates develops here is that Keynes' concern with anchoring economics in everyday language led him to 'formulate' an *ordinary language economics*. In the fifth chapter, he discusses Keynes' hermeneutic understanding of social theory and explains why conceptual adequacy supports the greater efficiency of ordinary language. Coates also raises a very relevant question about Weberian ideal types (models as samples) and the appropriate levels of reality for model building in the social sciences. In the sixth chapter, he pursues a more historical discussion, with the purpose of illuminating the 'striking' similarities in language and vagueness between the later views of Keynes and the Cambridge philosophers (in particular Wittgenstein). Finally, Coates returns to the Quinian framework of analysis and claims that the vagueness of ordinary language and common sense (as demonstrated by Wittgenstein and applied by Keynes) is in fact a valued property in *simplifying* the theory of complex systems. At the end of the book, he points out that a 'well-informed community of social scientists', based on common sense and ordinary language,must be considered as a 'powerful and efficient tool of discovery and communication'. He concludes that social sciences do not need a strict method of generation of knowledge grounded on formalisation to win their credence.

COMMON SENSE VS ORDINARY LANGUAGE

It must be emphasised that Coates' general argument hinges on the validity of the following logically independent elements:

(a) for all practical purposes, common sense and ordinary language can be considered as equivalent terms, with vagueness as their most distinguishing quality. In addition, Quine's contextualisation of common sense must be regarded as acceptable;
(b) the textual and biographical evidence of Wittgenstein's influence on Keynes must be persuasive;
(c) Keynes' acceptance of the notion of vagueness must be central to his 'ordinary language economies'.

The strength of Coates' general argument depends on the amount of evidence he is able to provide for each one of these conditions. In what follows we analyse the grounds for Coates' claims.

According to Coates, a paramount problem for the social sciences today is the existing gap between theory and practice due to the marginal role that the dominant philosophical paradigms attribute to vague concepts. Common sense, where vagueness is a distinctive feature, fills this gap between theory and practice, he argues. By common sense he understands not only 'the current state of theory', 'the practices of researching scholars', but 'our inevitable starting-points', our present conventions, our hypotheses about the world. In his opinion, the justification of common sense is given essentially by its critical function: we have only common sense in order to remake our conceptual world. Implicit here seems to be the idea that common-sense propositions are to be considered more 'certain' than mere philosophical propositions. In addition, he maintains that not only have other theories not offered enough to make us abandon our common-sense intuitions, but that there is not much sense in the idea that all our beliefs could be wrong, since we even need this inherited common-sense framework to be able to express doubt. These arguments, however cogent they may be, limit themselves to the justification of common-sense beliefs. They do not, however, aim to provide an explanation of the role of common sense in the progress of knowledge. As we discuss below, this issue should be examinedmore critically in Coates' work in order to have a clear idea of what he means by common sense.

In his 'Short history of common sense', two aspects should be noticed. The first concerns the practical equivalence he seems to adopt between common sense and ordinary language. There is no attempt to distinguish them. On the contrary, when discussing the work of Wittgenstein and Moore he claims that 'the case for common sense becomes closely tied to an analysis of ordinary language' (Coates 1996:21). Moreover, he maintains this practical equivalence throughout the book and uses it interchangeably. However, this distinction is of fundamental importance. In spite of similarities, common sense is not ordinary language. It

may be argued that, as stated by Bharadwaja (1977:65), 'the two are entirely different' notions, because whereas ordinary language is a method, a technique of philosophising, common sense is a theory, which expresses a certain world-view. This means that common sense is on a different and more abstract level than the notion of ordinary language. As Woozley (1953) has remarked, common-sense beliefs do not necessarily have to be formulated. As a matter of fact, we believe in things that we do not necessarily formulate. Of course, conceptualisation of the world is done through language, but the language is not the conceptualisation *per se*. Furthermore, assessing the truth of a commonsense belief is different from assessing whether ordinary language properly describes a situation or not, given that a person can use an ordinary expression ascriptively right but descriptively wrong. To a certain extent, Coates' position reminds us of N.Malcolm's (1942) claim that common sense is a disguised defence of ordinary language, which does not seem to be defensible because it ignores the distinction between ordinary language and common sense. The crucial question is, if these concepts cannot be used interchangeably, what are the consequences for Coates' argument? The first one is that all textual evidence raised by Coates should be examined with much more rigour as far as the differentiation between the two concepts is concerned. The second and more fundamental one is that this distinction can command a revision of the leading role that Coates assigns to the concept of vagueness, because while it is clear that vagueness is a central issue in Wittgenstein's later analysis, it does not have the same importance for common-sense philosophy. It does seem that both ordinary language and common sense reflect the same set of properties of cognition and environmental facts. Nevertheless, common sense is a more suitable framework than ordinary language for the description of the fundamental features of the world; the reasons were discussed above, namely that common sense is not a method but a world-view and that common sense may reflect beliefs that are not considered by ordinary language.

The second aspect which should be noted is that Coates, in his attempt to disentangle common sense from its past philosophical interpretations, embraces Quine's account of science and common sense. Broadly speaking,Quine identifies two elements in the progress of science: simplification and conservatism. More exactly, he sees the innovative feature of scientific progress reflected in the impulse to simplify, 'tempered by the drag of conservatism' represented here by common sense. As Coates observes (1996:31),

> If common sense has a claim to be heard its claim is nothing more than the counsel of conservatism in the choice of hypotheses. Listening to common sense is thus, as Quine says, both a 'counsel of laziness and a strategy of discovery'.

In spite of later claiming that Quine simplifies dramatically the theory of common sense, Coates remains throughout the book a believer in the conservative role of common sense. He accepts Quine's selection of factors but tries to show how they *may* depend not on Quine's ontology and canonical notation but on vagueness to promote efficacy in communication. Coates does not criticise the nature of the argument developed by Quine, merely his strong emphasis on the conservative role of common sense.

The crucial question to be asked here is why should common sense play a conservative role in the progress of science? Coates does not address this question directly, yet it is a question of fundamental importance. Commonsense beliefs are beliefs about matters of fact, about the world, from the perspective of individuals as social beings. Actually, common sense is a body of practical, empirical and social knowledge which guides our conduct in daily life; its distinguishing feature is its flexibility, meaning that one is not expected to continue upholding a hypothetical case when confronted with a concrete fact which contradicts it. Is this conservative? Well, we would suggest that it is not a 'yes or no' question, but a question of degree. If we accept that the degree of modification of our beliefs depends, among other things, on the progress of science, then it is evident that the 'degree' of conservatism of our common-sense beliefs should be assessed according to the speed of this progress. Therefore, we could claim for instance that during the Middle Ages, when systematic mechanisms for advancing knowledge did not exist, common sense would only undergo a process of slow evolution. However, this same line of thought does not apply to the case of contemporary advanced societies where science has contributed remarkably to the growth of knowledge. The upshot of this first group of comments is that Coates' contextualisation of common sense is not only defective, because it does not differentiate common sense from ordinary language, but it is also biased, because it embraces Quine's vision of science that provides a narrow account of common sense.

BIOGRAPHICAL EVIDENCE

The second group of comments turns to the question of how compelling is the biographical and textual evidence presented by Coates concerning Wittgenstein's influence on Keynes' work. Coates attempts to show that we can find some of Wittgenstein's later methodological insights by looking at Keynes' practice as an economist. He proposes to achieve it by following what he calls the 'counsel of prudence', which states that 'the place to look for valuable methodological insights' is in the successful practices of scientists.

What were Wittgenstein's core ideas and why are they important? According to Coates, many of Wittgenstein's later ideas, as much of late-Cambridge philosophy, are important because they pose a powerful criticism of the reductive nature of analytic philosophy that leads to an inversion of our notions of efficient

discourse. Wittgenstein criticised linguistic atomism arguing that not only do words not have a unitary meaning that can be established *a priori,* but that there is no such thing as a single logical structure behind our language. Rather, words have a 'cluster' of meanings, or to put it in another way, our everyday concepts are vague. Thus, in order to understand the meanings of the different words, we should use the notions of language-games and family resemblance, instead of attempting to specify defining essences. Meanings are displayed in the act of doing things, they are abilities and not mental states. Since concepts are vague, we cannot define their number of independent conditions of application, nor achieve clear identity conditions for them, and this undermines attempts at quantification and formalisation of concepts. However, according to Wittgenstein this is not a problem. On the contrary, concepts of everyday language (because they take into account vagueness) are more efficient at expressing meaning than formalised languages.

Coates follows two routes to correlate these insights of Wittgenstein with Keynes' analysis: the biographical and the textual. As far as the biographical evidence is concerned, Coates claims that

> It is evident that Keynes maintained an involvement in the intellectual life of Wittgenstein, was aware of the work he was doing, and considered it of importance to facilitate his continued publication. This makes it highly likely that Keynes' later analysis of vagueness and ordinary language derived to some extent from the work Wittgenstein was doing during the first half of the thirties.
>
> (1996:135)

Moreover, he supports the idea that in 1929, Keynes was perhaps Wittgenstein's closest friend when he returned to England. One feature of Wittgenstein's personality is added to complete the picture: his'single-mindedness' in his philosophising. Thus, he concludes (1996:130), 'it must be assumed that during these meetings throughout the early thirties Keynes and he [Wittgenstein] were discussing philosophy'.

This assumption is at odds with the account provided by one of Wittgenstein's most important biographers, Ray Monk. While Coates acknowledges it, he avoids discussing in depth the evidence provided by Monk. Monk argues that Ramsey, and not Keynes, was the closest friend and the 'most valued partner in philosophical discussion' Wittgenstein had after returning to Cambridge. There is textual evidence in the preface of Wittgenstein's *Investigations* that supports Monk's position. It is evident in Monk's biography that Wittgenstein had many friends during his life with whom he never discussed philosophy, which contradicts Coates' assumption that Wittgenstein discussed only philosophy with his friends. Furthermore, there is a vehement statement of Monk (1990:261 – 2), in direct antagonism to Coates' interpretation, that claims that

> Apart from Ramsey and Sraffa, Wittgenstein had little to do with the college dons at Cambridge. After the first few weeks his relations with Keynes were confined largely to business matters, and although Keynes became an invaluable ally whenever Wittgenstein needed anything sorted out with the authorities, he was not a close friend. This, one gathers, was a role that Keynes was quite happy to fit into; being Wittgenstein's *friend* demanded more time and energy than he was able, or prepared, to give.

Monk documents Wittgenstein's legitimate anxieties about Keynes seeing himself more as a benefactor than as a friend. The expression Monk chooses in order to best qualify Keynes' role in their relationship is that of a 'friendly benefactor'. It is difficult to ignore this argument. Keynes' diverse interests, as observed by Skidelsky (1992:388) in 'writing, broadcasting, corresponding, teaching, bursarial work, committee work, City work, ballet work', etc, kept him so busy as to not leave him much time and attention for Wittgenstein's philosophical conundrums. Furthermore, it might be too much to expect that Keynes would understand Wittgenstein's new theories in a few hours while full-time top professional philosophers took much time and effort to accomplish this task. It is a well known fact that Keynes, who was acquainted with the analytic tradition, struggled to understand the *Tractatus*. Why should he understand better these new ideas when his attention was divided among so many interests?

What could be argued on behalf of Coates' biographical interpretation? Perhaps that Keynes was acquainted with some key Wittgensteinian concepts and that he employed them in his theories, not because he had understood their philosophical significance, but simply because they made sense in a different context. Is this not exactly the practical meaning of thelanguage-game technique? Does it not mean that there is no need for justification for a game, once one can *play* it? We return to this important question below when discussing Coates' arguments about Keynes' economics. For the moment, we can only conclude that there is no unequivocal biographical evidence regarding the relationship between Keynes and Wittgenstein.

As far as the textual evidence is concerned, Coates emphasises that both Keynes and Wittgenstein focused their analyses on the property of vagueness and on the criticism of formalisation of concepts. The idea is that Keynes would have found in Wittgenstein's later philosophy 'a logical vindication of his earlier reservations' to Russell's reductionist symbolism and Moore's intuitionism. Consequently, Coates holds that vagueness is manifested as a central concept in Keynes' economics in the criticisms of:

1 the use of reductive symbolic analyses;
2 the possibility of direct knowledge, and
3 the atomic hypothesis.

He argues that much of Keynes' criticisms of the classical economists and his analysis of expectations and uncertainty are an outcome of this influence.

TEXTUAL EVIDENCE

It must be noted that Coates attributes to Keynes an almost exclusively philosophical logic. He hardly discusses the consequences (reactions) of Keynes thinking as an economist, a businessman, a man able to be influenced not only by philosophical arguments but to the 'ordinary facts of life'. However, a comprehensive view of Keynes' thought must take into account these practical elements. Is it really relevant to claim that Keynes knew the philosophical grounds of common sense and ordinary language? Is it not much more important that he was able to use his common sense? Does it actually matter if Keynes' common sense, translated or not into equations, institutional constraints, etc, were to be based on his 'unanalysable intuitions' or on ordinary language concepts?

It should be borne in mind that in Moore's analysis the notion of intuition was the main one responsible for providing the grounds through which people could have direct access to reality. Later, Cambridge philosophy distrusted this reliance on the atomist model of correspondence presupposed by Moore's phenomenalism and on the process of direct apprehension of reality. Moore's common-sense realism, however, was retained. With Wittgenstein came the idea that meanings depend upon language-games and the circumstances in which they are used; that an individual's understandingof reality is not achieved through an immediate and unmistakable intuition, but through a sort of interpersonal social intuition given by the social linguistic framework. However, it must be noticed that whatever the case, the basic function of 'intuition' is the same, namely to provide grounds for our common-sense beliefs. Thus we are willing to suggest here that Keynes did not have a 'logical vindication of his earlier reservations' concerning Moore's and Russell's philosophy since he could operate his common-sense apparatus of thought *as an economist*. Coates perhaps 'misses the target' when he focuses his discussion of Keynes' economics on ordinary language. In any case, we would suggest that Coates overemphasises the direct influence that Wittgenstein's philosophy had on Keynes, to the detriment of the idea that Keynes' philosophy was transformed by development of his economics. Indeed, this crucial element leads us to Coates' last set of comments, namely, that Keynes would have formulated an 'ordinary language economies'.

To follow an ordinary language economics means that, given the complexity of economic systems, 'it is best not to stray too far from concepts and mental constructs with which we have an immediate understanding' in order to avoid confusions. To a certain extent, this interpretation is in accordance with Quine's emphasis on the conservative role of common sense (more exactly, the conservative role of ordinary language). It is also reflected in Coates' claim that (1996:105) 'There is indeed something to the idea that common sense is

isomorphic with the subject-matter of the social sciences' (emphasis added). However, this is a caricature of common sense, for many of our common-sense beliefs deal with counter-intuitive and 'hidden' phenomena that do not have any isomorphic relation to reality. Coates, following this interpretative framework, argues that Keynes placed 'greater weight on the virtue of conservatism in social theory' (1996:38). According to him, this is manifested in three ways. First, in the lengthy treatment of the definitions in *The General Theory,* such that they correspond with actual usage. Second, in the choice of its units in the measurement of the system as a whole, where Keynes criticised the 'mock precision' of quantitative analysis. Third, in Keynes' analysis of the fragility of expectations, where he applied the concept of vagueness. In these three instances Keynes' concern lay with 'anchoring economic analysis in everyday language' in order to achieve a more efficient discourse for handling the complexity of economic relations than that provided by formalised languages. Coates recognises that some of the above elements can be found in Marshall's comments and therefore claims that (1996:93) 'Thus there was something of a precedent for ordinary language economics in Cambridge'. Ultimately, what Coates is saying is that Keynes was a sort of incipient ordinary language scientist, who after being in touch with Wittgenstein's ideas became an ordinary language scientist in his plenitude.

In order to justify his interpretation, Coates mentions more textual evidence concerning

1 Keynes' doubts about formalisation,
2 the philosophical use of the word 'vague' and the discussion of the question of precision in economics, in Keynes' 1932 – 3 lectures,
3 the almost absence of equations and symbolism in *The General Theory* in comparison with the *Treatise on Money,*
4 Keynes' dismissal of the superiority of Joan Robinson's symbolic method,
5 Keynes' correspondence with Harrod, where he discusses the 'problem of symbolism', and
6 the similarity between Keynes' use of models as samples and Wittgenstein's ideal constructs.

This leads Coates to interpret Keynes' economics in the light of hermeneutic lines pointing out that (1996:100) 'A social theory must be "adequate" to its subject, it must be capable of being stated in terms which the subjects could understand, and with which they could identify'. He does not, however, provide enough textual evidence to support the 'familiar hermeneutic case' made by Keynes. Be that as it may, some of his quotations and statements such as (1996: 101) 'Keynes' was "an analysis which is endeavouring to keep as close as it can to the actual facts of business calculation"' seem to unduly interpret common sense as ordinary language. His defence of Keynes' belief in explaining 'the

familiar in terms of the familiar' suffers from the same lack of distinction between common sense and ordinary language. In fact, many of these quotations reflect more Keynes' concern with an appropriate modelling of real-world facts (a realistic world-view) than his defence of an 'easy-to-read' economics.

Towards the end of the book, Coates assesses the influences of Sraffa, Ramsey and Marshall on Keynes, and does not seem very convinced of his earlier claims about the supremacy of 'the striking similarities' between Keynes and Wittgenstein. It even seems that he is recapitulating his position when he recognises that he does not (1996:144) 'think that Keynes' change in method was simply the result of his incorporating into his economics recent advances in Cambridge philosophy'. In spite of the apologetic tone he does not give an alternative explanation for his previous claims.

Does Coates provide enough reasons to justify his claims that Keynes formulated an ordinary language economics? It is difficult to achieve a final verdict on this question because the many reasons he provides are based on not very compelling textual evidence. At first, several features of Keynes' economic analysis give the impression of being explained through the notion of ordinary language, such as his critical attitude towards formalisation, the role of expectations as a social phenomenon, the importance of defining theinstitutional and historical set of analysis, and his commitment to realism in economics. Subsequently, when we examine thoroughly the evidence provided by Coates, we realise that most of it refers not to ordinary language but to common sense, and that specifically in *The General Theory:*

1 The choice of concepts and definitions may be partly explained by Keynes' embrace of the latest advances in economic science during the late 1920s and early 1930s and by previous public misunderstandings of his message, and not, as Coates suggests, by Keynes' decision to *formulate* an ordinary language economics. For instance, the concept of effective demand, as the main cause of unemployment, was used in 1925 by Oswald Mosley, and was not an 'ordinary language invention' as claimed by Coates. Richard Kahn's 1931 multiplier was first associated with the consumption function by Jens Warming in 1932, and later incorporated into Keynes' analysis because it was a tool for explaining the logic of the great depression. Keynes' appropriation of Sraffa's construct of commodity rates of interest for Chapter 17 of *The General Theory* pursued more theoretical aims. Another example comes from a seminar in June 1935 at LSE where the main objections raised against *The General Theory* were about its definitions of income and investment. It follows that Keynes' emphasis on definitions *may* have been triggered by external causes, and not by his ordinary language commitments, as proposed by Coates. The question of the use of vague (or precise) concepts is one to be settled by the problem at hand and not by *a priori* philosophical adherence to linguistic conceptions.

2 The choice of units follows a different pattern from the one claimed by Coates. For instance, he points out (1996:83) that Keynes

> found the treatment of units of quantity in other economists unsatisfactory *because of* their attempts to quantify essentially 'vague concepts, such as the quantity of output as a whole, the quantity of capital equipment as a whole and the general level of prices'.
>
> (emphasis added)

This is a misinterpretation of the quotation. What Keynes said (1973: 43) is that when dealing with the economic system as a whole it was better to use only two fundamental units of quantity, namely money and labour, because this was a more satisfactory system of units, and not because the other variables mentioned above have no function to perform. Keynes' point here is against 'the use of units of particular outputs and equipments' for the analyses of the economic system as a whole. It reflects an absence of adequate national accounts, and not a defence of vague concepts because they reflect ordinary language. As a matter of fact, the social policy derived from *The General Theory* calledfor quantification of the main aggregates, such as consumption, expenditure, income and investment. It could even be argued that the static nature of the equilibrium method used in *The General Theory* (in comparison to the method used in the *Treatise*) was due to the fact that 'disequilibrium analysis failed to yield *precise* concepts and tools of policy' (Skidelsky 1992:601, emphasis added).

3 Much of what Keynes meant by vagueness probably came 'directly' either from Marshall's teachings or from Keynes' own understanding of the historical reality of the 1930s. For instance, Marshall in his *Principles* (1920: 25) argues that measurements in economics are 'seldom exact' and that there are no precise laws in economics (*ibid.:* 26) 'For the actions of men are so various and uncertain, that the best statement of tendencies, which we can make in a science of human conduct, must needs be inexact and faulty'. Moreover, Marshall not only defends explicitly the ordinary language method (subordinated to common sense—'the ultimate arbiter in every practical problem'), but does it from a Sraffian-anthropological perspective (that influenced Wittgenstein so much) when he argues that (*ibid.:* 43) 'In common use almost every word has many shades of meaning and therefore needs to be interpreted by the context'. In addition to this evidence we should remember the long-established Cambridge maxim that 'it is all in Marshall', and conclude that the case for Marshall's influence is much more serious than Coates admits in his book. We return to this issue in more detail below.

COMMON SENSE ECONOMICS

It must be noted that in his Ph.D. Thesis on 'Ordinary language economics: Keynes and the Cambridge philosophers', submitted in 1990 to the University of Cambridge, on which the book is based, Coates not only approached more precisely (and extensively) Marshall's influence, but put much more emphasis on the role of common sense as Keynes' guiding principle of analysis. That is why we qualified above the relation between the thesis and the book as ambivalent. On the one hand, Coates changed the title of his work in order to emphasise the role of common sense, and even added a chapter with a review of the history of common sense. On the other hand, he omitted and modified previous arguments he had developed which supported common sense. As a consequence, Coates' interpretation of Keynes' criticism of classical economics was seriously impaired. Whereas in the thesis he showed that (1990:106) 'Keynes tried to call economists back from accepting that conclusions drawn with the aid of methods that somehow look like those of the natural sciences are more securely founded than our common-sense understandings of the world', in the book (1996:88) he prefers to identify the source of disagreement between Keynes and the 'classics' as 'the issue of the degree of precision'. For people who read his thesis and later the book, the progress in his line of thought could well be puzzling.

The comparison between the thesis and the book might inspire a constructive critical reaction, perhaps, if we concede that vagueness in Keynes might be a result of his observation of particular historical and institutional economic conditions, under the influence of Marshall's commitment to common sense according to which economics must 'deal with man as he is' and 'go hand in hand with facts'. Put differently, if we consider vagueness not as a philosophical/methodological demand but as a stylised representation of the particular experiences of his time, then we are willing to suggest that Keynes would have *followed* the Cambridge tradition of a *common-sense economics*. The theory Keynes inherited, built on the economic conditions of nineteenth-century England, was not able to explain the collapse of output and employment seen during the depression of 1920 – 2 and the Great Depression of 1929 – 32. Keynes' experience of monetary disorders (especially in 1924 – 5), built from his understanding of the Treasury's and the Bank of England's *laissez-faire* policies, made him realise how economists and economics assumed away the basic features of the social system in which they lived. The context of the great output slump of the early 1930s consisted of high sustained unemployment, active union behaviour, increasing influence of the banking system in determining money supply, supremacy of the joint-stock company, volatility of the financial markets and other elements inadequately portrayed by neoclassical economics. Hence, it seems reasonable to agree with Skidelsky (1992:87) that

> it was the shift from stable to unstable social structures—or more generally the increase in historical turbulence—which caused Keynes to theorise the

> economic problem in terms of irreducible uncertainty. If this interpretation is true, uncertainty itself is a historical variable, its scope, like that of probability, depending on time and place.

However contentious this claim might sound, it illustrates Keynes' willingness to theorize according to the realities of his time. Keynes' criticisms of neoclassical theory do not just reflect his disagreement with the degree of precision of the methods used in economics, but first and foremost they manifest different world-views: academicism versus common sense.

More exactly, the clash between neoclassical economists and Keynes was about building economic theory from conceptual notions (based upon concepts by postulation), such as the barter economy, or building it from reality as we conceive it (based on concepts by intuition). This is consistentwith Harcourt's (1996:98) distinction between the two basic and different approaches to economics, one of which is the axiomatic and 'the other [which] starts by observing behaviour, institutions, "stylised facts" and then constructs simple models incorporating the essence of the observations in order to try to explain the original observation[s] *et al.*', which we suggest be properly named *the common-sense approach to economics*. To re-emphasise: neoclassical economists clung to a greater extent to their unalterable prewar views about the nature of economic systems than, as suggested by Coates in his book, to their excessively precise methods of analysis.

As far as the methodology of economics is concerned, Coates' arguments are out of focus. His emphasis on the question of vagueness obfuscates the more important question of common sense as a guiding principle of analysis distinct from the method of ordinary language. He seems to ignore the possibility of coordinating a common-sense approach to economics with methods other than those uniquely based on ordinary language. In the same way, Keynes' method should be assessed according to criteria wider than those proposed by Coates. To restrict Keynes' methodological contribution to the logic of ordinary language is to impoverish it.

However, we should not let these criticisms befog the extreme relevance of Coates' argument. The question he addresses about the gap between the theory and practice of social sciences is the most imperative problem to be settled in economics since the advent of the formalist revolution due to the spread of axiomatic methods. Moreover, the solution he proposes is on broad terms on the right track. The theoretical elements he provides, if reorganised, can be used to make a very solid case for what we suggest should be called *common-sense economics*. Once Keynes' philosophical development be contextualised under this perspective, we can transcend the usually self-centred and hermetic study of Keynes' thought and proceed to a wider investigation of common sense as a theorising engine—which certainly would include the works of Marshall, Sraffa, Kaldor, Joan Robinson and many others.

REFERENCES

Bharadwaja, S. (1977) *Philosophy of Common Sense,* New Delhi: National Publishing House.

Coates, J. (1990) 'Ordinary language economics: Keynes and the Cambridge philosophers', Ph.D. thesis, University of Cambridge, mimeo.

——(1996) *The Claims of Common Sense: Moore, Wittgenstein, Keynes and the Social Sciences,* Cambridge: Cambridge University Press.

Harcourt, G.C. (1996) 'How I do economies', unpublished manuscript, February.

Keynes, J.M. (1973) *The General Theory of Employment, Interest and Money,* in *The Collected Writings of John Maynard Keynes,* vol. VII, ed. D.E.Moggridge for the Royal Economic Society, London: Macmillan.

Malcolm, N. (1942) 'Moore and ordinary language', in P.Schilpp (ed.) *The Philosophy of G.E.Moore,* New York: Tudor Publishing, 343 – 68.

Marshall, A. (1920) *Principles of Economics,* 8th edn, London: Macmillan.

Monk, R. (1990) *Ludwig Wittgenstein: The Duty of Genius,* London: Vintage Books.

Quine, W. (1960) *Word and Object,* Cambridge MA: MIT Press.

Skidelsky, R. (1992) *John Maynard Keynes: The Economist as Saviour, 1920 – 37,* London: Macmillan.

Woozley, A. (1953) 'Ordinary language and common sense', *Mind,* vol. LXII, 301 – 12.

12
SCIENTIFIC COMMUNITIES, LANGUAGE AND THE KEYNESIAN REVOLUTION

Mario Garcia

The impact of Keynes on economics was so great that it has often been referred to as the 'Keynesian revolution', and there has been some discussion about whether it was a scientific revolution in Kuhn's sense. The aim of this essay is to re-examine this problem in the light of Kuhn's later discussions on paradigms; particular attention will be paid to the identification of the scientific community and the role of language in the revolution. The problems of applying the concept of paradigm, in the fuzzy sense used in *The Structure of Scientific Revolutions,* to the Keynesian case are explained in the first section; the three possible scientific communities are then examined; the role of language in them is studied in the third section. The final section contains some concluding remarks.

SOME PROBLEMS OF A KEYNESIAN PARADIGM

In his book on scientific revolutions, Thomas Kuhn (1962) characterised science as a discontinuous process. In normal times, scientists do not criticise the fundamentals of their discipline and instead devote themselves to the puzzle-solving activity of applying and developing those fundamentals. However, from time to time some intractable puzzles appear that lead some scientists to question the fundamentals. A new view appears that solves the puzzle but is incommensurable with the former view and, therefore, neither of them can show its advantage over the other. Kuhn called each of these incommensurable views a paradigm. After a time of crisis during which some members of the community convert to the new paradigm, textbooks are written that fill the gap and give the idea that the new paradigm evolved from the last one in a continuous process. And then a new period of normal science begins, which has its own problems and criteria for their solution.

Can there be paradigms in economics?

The application of Kuhn's theory to social sciences and particularly to economics is problematic (Katouzian 1980; Blaug 1980; Barnes 1982).[1] First, the coexistence of radically different views in the discipline suggests that economic

paradigms do not die (Brofenbrenner 1971). Second, in contrast with the immutability of the natural world, the historical nature of social and economic reality implies that a paradigm makes sense only within a rather specific set of institutions (Kunin and Weaver 1971:395). These arguments, plus the imprecision of Kuhn's definition of a paradigm (Stigler 1969), led some authors to reject the use of the concept in economics in favour of a dialectical change (Brofenbrenner 1971; Karsten 1973).

Other authors accepted the relevance of the concept but applied it cautiously, particularly because there is no agreement about which of the revolutions in economic thought have involved a change of paradigm (cf. Deane 1984; Coats 1969). Coats (1969), for example, argued that economic paradigms are less precise and less liable to falsification, that is, there are no crucial experiments. It also seems that ethical problems and economic policy play a special role in triggering revolutions in economics (Dillard 1978).

Is economics a mature science?

It is, of course, possible that all these problems arise from using a saw to do a haircut. Kuhn's theory was developed to explain natural sciences, not proto-sciences like the arts, philosophy or social sciences. His purpose was to '[c]onfine attention first to fields which aim to explain in detail some range of natural phenomena. (If…my further description fits theology and bank robbery as well, no problems are thereby created.)' (Kuhn 1970:245). A field is mature, he continues, when provided with theory and technique that bring about consistent predictions over a specific subject and a way to improve these predictions *(ibid)*.

Whether economics makes this kind of prediction is a complex problem that will not be discussed here. It is enough to remember that event regularities seem rarely to occur in the social realm, which makes econometric forecasts unsuccessful outside the period for which the models were initially constructed (Lawson 1997:36, 71 – 2). On the other hand, prediction is not a necessary condition for a science. Kuhn (1974) himself has provided an example of a science with paradigm that does not make predictions, namely, taxonomy.

A second criterion for maturity is the isolation of the community, in the sense that the sole audience for its members is the community itself. This, again, is in contrast with social sciences 'that still aim to communicate with and persuade an audience larger than their own profession' (Kuhn 1970:254). Although he argues that esoterism is accomplished bymodern economics *(ibid.),* ideology and the features of economic policy have a role in economics that cannot be dismissed in the appraisal of the second criterion.

A first try

The most extensive attempt to explain the Keynesian revolution after Kuhn's work was made by Ghanshyam Mehta.[2] In a Ph.D. dissertation at Berkeley, accepted in 1971 and published in 1977, he argued that *A Treatise on Money* (1930) was the beginning of a new paradigm; that this book already contained discussions, such as the determination of levels of output, the multiplier or the rejection of Say's Law, that would appear again in *The General Theory*. The discussions between Keynes, Robertson and Hayek would show that they had different paradigms.

The idea that everything was in *A Treatise on Money* is difficult to sustain in the light of *The Collected Writings,* Skidelsky's (1992) biography, Harcourt and Riach (1997), Patinkin (1976), Patinkin and Leith (1978) and other studies. But it is also clear that certain concepts, such as the user cost or the determination of spot and future prices, had a root in *A Treatise on Money* (Davidson 1987; Kregel 1997). Indeed, some of the continuities were more evident for Keynes' contemporaries than for us. By the time Keynes introduced the term 'effective demand' into the drafts of *The General Theory,* for example, McCracken (1933) had already linked Keynes with Malthus against Say and Ricardo in the use of 'effective demand' and the rejection of Say's Law. Another example is Adarkar's (1935) use of Keynes' backwardation theory in the determination of Sraffa's natural rates of interest, a procedure that Keynes would also use in Chapter 17 of *The General Theory* (see also Lawlor 1994). Better known is Joan Robinson's 1933 article linking money with the theory of output.

But even if *A Treatise on Money* anticipated *The General Theory,* showing that the contemporaries had difficulties in accepting the theory is not enough to prove that there was a change in paradigm.

Mehta's handicap was the state of knowledge about the 1930s at the time he wrote his dissertation, which has considerably improved during the last twenty years. His problems were the result of the weakness of a non-historical approach to the identification of paradigms (Kuhn 1970:251). But there are also problems with the concept of a paradigm itself. It is possible to apply here Kuhn's later thoughts about paradigms and scientific revolutions (1970, 1974). As an answer to his critics, he described how the process of studying a scientific revolution would be. The two main features were the identification of the scientific community relevant for the case studied, and the creation of a new language (the paradigm in the strict sense) that allowed scientists to see new phenomena. In the new approach, the assumption ofan esoteric community should be relaxed in order to identify the scientific community.

THE SCIENTIFIC COMMUNITY

There is no definite way to identify the relevant community, but three candidates can be taken into account:

1 the implicit reader of the book;
2 the community of theoretical economists that actually read it; and
3 the second group plus the policy makers that accepted the 'New economies', as Keynes' doctrine was soon known.

The implicit reader of the General Theory

This book is chiefly addressed to my fellow economists'. The first lines of the book make it clear that its reader, unlike the one of *A Treatise on Money,* is a member of an esoteric community. The general public is welcome to the debate only as an eavesdropper. The controversial attitude is stressed by the repeated contrast between 'I' and 'my fellow economists' in the preface to the first edition.

These fellow economists, we are told, believe in the Classical Theory, that is, they belong to the tradition of Ricardo and Marshall. The dominant role of the Marshallian orthodoxy in England is also the initial topic of the prefaces to the foreign editions of the book. Thus the reader is a Marshallian economist, not a member of other schools, which fits with the then-predominant Cambridge view that economics was made in Cambridge and that 'it is all in Marshall' (Turner 1989:15; Rymes 1989:18). This may explain the role of Kahn in the writing of the book, for it means that Keynes used him to discover to what extent a sympathetic intelligent reader brought up on Marshall would understand Keynes' arguments.[3] As Paul Davidson suggested (King 1994), if Kahn did not accept a point, Keynes could be sure that other readers would not either. In this case, when Keynes was still sure about his own argument, he would relegate it to a footnote or an appendix (as in the case of the user cost) waiting for the moment the reader would discover the problem. This was a custom Keynes learnt from Marshall (Fouraker 1958), to whom we shall return. Thus Kahn can be seen, to a great extent, as Keynes' model for the implicit reader of *The General Theory*.

It could be argued that the implicit reader might be an Austrian, because the passage on the divergence of opinions between economists that has 'almost destroyed the practical influence of economic theory' (Keynes 1936: xxiii) is probably an allusion to the committee of economists that worked parallel to the Economic Advisory Council in 1930 and to the failure of itsmembers to overcome their differences and propose a remedy for recovery. Among the members of this committee, the opposition to Keynes' ideas came from Pigou, Henderson and mainly Lionel Robbins, from LSE, whose theories derived from Austrian economics (Skidelsky 1992:368). The way Hayek is treated in *The General Theory,* mentioned sometimes indirectly as a 'wild duck' belonging to a certain muddling 'neo-classical school' (Keynes 1936:183) would be intended to stimulate controversy. However, the idea of an Austrian implicit reader is valid only to the extent that the Austrians had a strong Marshallian background in theory and style.

If the idea of a Marshallian implicit reader is accurate, it would help to explain the different reception of the book between those who had been exposed to the Marshallian style and those who had not, e.g. those in America.

Theoretical economists

By the 1930s, the professionalisation of economics had been accomplished in Britain and the USA, and there existed organisations of economists (see Coats 1964, 1985; Coats and Coats 1973). The *Economic Journal,* the periodical of the Royal Economic Society, 'never attempted to cater for the general public' (Coats and Coats 1973:165). As its editor, Keynes played an influential role in the profession (Coats 1991:102ff). Those people in the profession working on theoretical issues would form the second candidate for the scientific community.

Due to its controversial style, *The General Theory* was initially received with irritation and criticism, and several of these first reviewers claimed that the book was not 'a revolution in fundamental economic theory' (Harrod, quoted by Harris 1947; Skidelsky 1992:574 – 9). But this effect was surely good for the controversy that Keynes was looking for: 'because it is only out of the controversy that will arise that what I am saying will get understood' (Keynes 1973a:548).

The controversy had more effect upon the young economists of a limited number of universities: Cambridge, LSE, Oxford, Harvard and Chicago (Moggridge 1995). A small group of theoretical economists, especially Kahn and Joan Robinson, worked hard to spread the word. Abba Lerner, for example, who was already a first-rate theoretician, spent an unusual amount of time writing articles divulging the principles of *The General Theory* (Scitovsky 1984). Lorie Tarshis, Robert Bryce and Walter Salant carried the message to Harvard where Hansen and Samuelson would take it, although they were not converted as were those in England (Turner 1989:55, 59; Colander and Landreth 1996:72). Reddaway, Champernowne, Harrod, Meade, Hicks and Lange formalised the theory, making it more accessible for those economists more inclined to formal methods (although the meaning changed in this translation; Young 1987:30 – 59; Darity and Young1995). In three of these cases, the writer had in mind the Econometric Society, a growing group within the profession.

The tendency to formalisation can be seen in the increase of works on the quantity theory during the 1920s. These works presented improvements in the statistical techniques together with an increasing poverty of theory (Humphrey 1973). The resulting degenerated form of the Fisherian theory did not resist the impact of the Keynesian revolution and the studies moved towards the estimation of the Cambridge equation (Burghardt 1975), in which the velocity could be considered as the inverse of the demand for money.

There was another tendency that worked alongside the Keynesian revolution. Gathering and estimation of data improved during this period, and later as a

result of the war. Kuznets was then collecting data that would be applied to *The General Theory* (e.g. the consumption function). In England, Keynes played an important role in pushing the government to collect statistics if the war was to be won (Harrod 1974:14). The creation of the Department of Applied Economics at Cambridge was in the same line of empirical work that was dear to Marshall (Coats 1967:714).

If the scientific community is defined as those members of the profession with a theoretical inclination, the interaction between Keynes' theory and the tendencies among this group gains importance as a factor explaining the success of the revolution.[4] This approach, however, faces the problem that the line between theoretical and applied economists was not always clear. Besides, the rejection of the revolution in Chicago was related to its application. They considered the book mainly as a short-run theory, the policy implications of which were not new, since government expenditure as a remedy for unemployment had been already deduced from the quantity theory and proposed by the Chicago economists during the early 1930s (Davis 1968; Patinkin 1981). This aspect leads us to the third group.

The policy-makers

In the absence of experiments in economics, the apparent success or failure of an economic policy, can play an important role in the rejection of a theory, even though economic policy cannot falsify it. In addition, the changing character of the economic world means that theories have to cope with the new aspects of the reality, the results of policy being the most evident of those aspects. If this is the case, the appropriate scientific community should include policy-makers.

This community included not only economists in Cambridge and the UK but also in the US.[5] The impact of *The General Theory* upon this group in Washington was channelled through Alvin Hansen's seminar at Harvard's Graduate School of Public Administration. But the policy-makers who attended the seminar were 'neither converts nor theoretical Keynesians'(Turner 1989:59; Colander and Landreth 1996:200). It was the recession of 1937 that convinced some economists, Hansen among them, to accept Keynes (Barber 1996:111; Stein 1969:165).

The young economists working in the New Deal government were seeking a theoretical support for the policies they were already promoting, but were also too busy with the practical aspects of economic policy to fully study the implications of *The General Theory* (Colander and Landreth 1996: 224). Although influenced by Keynes before *The General Theory,* the intellectual leader of this group, Lauchlin Currie, was highly sceptical about several aspects of *The General Theory,* from the definition of saving and investment to the utility of the multiplier as an empirical tool (Sandilands 1990:36, 84 – 7; Currie 1938).

The impact of Keynesianism on economic policy has been described several times (e.g. Klein 1968; Stein 1969; Booth 1983; Colander and Coats 1989). Enough to say that, at this level, the theoretical implications of *The General Theory* were easy to forget; that the book would become an unread classic, being replaced by Hicks, Hansen, and later, Samuelson. The ISLM model, in a period of sustained growth, gave the impression that economic policy was something easy to do.

After World War II (Colander and Landreth 1996:72, 84) a new generation found it easier to accept a surface Keynesianism that was promoted by the textbooks that followed Samuelson (1947) and preserved the Classical view in the chapters on microeconomics (Colander and Landreth 1996: 20 – 4). Samuelson himself did not convert but merely stopped resisting Keynes' ideas (Colander and Landreth 1996:159 – 60) and the subsequent editions of his book have less and less Keynesian elements (Pearce and Hoover 1995).

LANGUAGE

In Cambridge

Language had been a long-time preoccupation at Cambridge. Marshall and Keynes preferred to use ordinary language and relegated mathematics to appendices or footnotes. This allowed them to avoid the false precision of formalism and to reach a broader audience.[6] When attacking a problem, they would use all the resources available, including mathematics; once it was solved, they would make an assumption that allowed the reader to avoid all the difficulties that they had before and translate the argument into ordinary language. Afterwards, perhaps in a footnote, they would discuss the effect of dropping assumptions and all the possible criticisms. Hence the reader would have an initial understanding of the problem, overlooking the obscure passages. Once the reader had assimilated the fundamental principleand discovered the difficulties of the argument, the passages that were previously obscure would make sense (Fouraker 1958).

This technique had the advantage of making the book more accessible, without the need of handbooks or exercises (exemplars), but readers from other stylistic traditions would easily become impatient and would not go back to the book when they found a problem. Leontief (1937) for example, was annoyed by this seemingly tautological style, for which he coined the term of 'implicit theorising'.

However, those raised in the technique would be conscious of the use of language. This pushed in the same direction as the philosophical background of Cambridge, first with Moore's emphasis on common sense and ordinary language (Malcolm 1952) and later with Wittgenstein's discussion of vagueness (Coates 1997).

This language-consciousness explains Keynes' stress on units and definitions which Bradford and Harcourt (1997) have discussed. A complete chapter of *The General Theory* was devoted to the meaning of the word 'saving' and a great deal of his 1937 controversy with Ohlin was about the definitions of ex-ante and ex-post saving, and finance (Keynes 1973b: 201 – 23, 229 – 33). Definitions had also a role in Keynes' discussion with Robertson (Keynes 1973a:327 – 30, 493 – 507).

The successful terms

The expansion of Keynes' ideas can be seen in the victory of the terms he used. First, as mentioned before, even the first unsympathetic reviewers used the word revolution, even if only to deny its existence. The rhetoric of controversial revolt also raised the issue of 'Keynes versus the Classics' among the young (Moggridge 1986:358). The term 'aggregate demand' implied that the apparatus of supply and demand was not being used in the traditional way, but for the whole economy. Thus, this notion would pave the way for Keynes' division of economics between 'the theory of the individual industry or firm…and the theory of output and employment *as a whole*' (Keynes 1936:293, original italics).

Some of the Keynesian terms were conserved definitely in the National Accounts, but not all the terms succeeded. Since Keynes was read through the works of Hicks and Hansen, those concepts that created problems for the simultaneous equation approach (e.g. the finance motive) were forgotten or replaced by others that could be treated more easily in mathematical terms (e.g. risk instead of uncertainty in Tobin 1956). Two successful expressions, money demand and liquidity preference, induced the idea that people would demand an asset with no return. But when translated into the language of the ISLM, it became a stable relationship between the amount of money and the rate of interest. The stability of the relation was needed in order toexpress liquidity preference as a mathematical function, and was later reinforced by the belief that the function could be econometrically tested.

The New Dealers' semantics

In the US, the economists of the New Deal experienced the problem of finding the appropriate vocabulary and metaphors to support their policies. One example is the use of the term 'pump priming' to describe the mechanism set in motion by government expenditure. The idea was successful because it was familiar to the general public: 'so many people then used hand pumps and their experience was that pouring in a little bit of water for a short time started a copious and sustained flow in a normal fashion' (Keyserling 1987). The use of this metaphor had theoretical implications because it led to the conservative idea that intervention was needed only at the beginning, for later the market mechanism would work

by itself. Because of its semantic appeal, the expression continued being used until the beginning of World War II despite the fact that as early as 1935, its main advocator, Lauchlin Currie, recognised it as misleading for 'the government might have to keep pumping and pumping for some time' (Sandilands 1990:73).

One of the main problems for the New Dealers was that

> not only did the word 'spending' have bad connotations—being commonly linked with such adjectives as 'loose', 'reckless', or 'wild'—but there was much confusion about its economic effects. Different types of spending were evaluated according to…[different] implications. Few people saw that there is a common element in all of them.
>
> (Sweezy 1972:116)

> In 1934, the Treasury 'began calculating a series which was first called The Pump-Priming Deficit', then later 'Federal Income Increasing Expenditures' and finally the 'Net Contribution of the Federal Government to National Buying Power'. This was both a technical improvement…and even more important a semantic triumph of the first magnitude. It brought out the common element in all the government's fiscal operations. No one used to thinking in terms of the net contribution could advocate promoting recovery by increasing public works spending while at the same time cutting government salaries and raising tax rates.
>
> (Sweezy 1972:118)

This second quotation is the best example of a paradigm in the revised sense of Kuhn (1970, 1974). The new language allowed people to see different things that, for someone outside the paradigm, were invisible or non-existent,or puzzling and irritating (Bradford and Harcourt 1997:107 – 8). Nevertheless, it should be kept in mind that language interacted with ideology. Part of the success of Samuelson's textbook (1947) over Tarshis's (1947) lay in its use of graphs, diagrams and mathematics. Formal language created a 'scientific' aura that allowed it to avoid the rejection that Tarshis's book found for its ideological implications (Colander and Landreth 1996:23, 66 – 9).

THE FUZZY PARADIGM

In *The Structure of Scientific Revolutions,* the term paradigm was used in several different senses. This was a polemic feature (Masterman 1970) and Kuhn himself recognised that the term was used beyond its initial objective (Kuhn 1970, 1974). In trying to assess the relevance of the concept for economics, Phyllis Deane (1984) has stressed three meanings of the term: a metaphysical one, in which the paradigm is a world-view; a sociological one, as a set of

scientific habits and procedures; and a set of a problem-solution and the analytical toolkit currently accepted by a scientific discipline. Although she found that there were changes in all three senses, she concluded that this was an unsuccessful revolution because those changes depended more on the new conditions of the real world than on the ability of the intellectual community of economists to change their habits of thought. As a result, when the real world changed again in the 1970s, the old theories crept back (Deane 1984:10, 11).

Nevertheless, it is worth reconsidering the question in the light of the previous remarks on scientific communities and language. Did the Keynesian revolution change the world vision? The answer to this question depends on what is meant by 'world vision'.[7] As related to economic policy and the role of the state, the answer is yes. *The General Theory* gave the theoretical support for state intervention in a way that is still important. From a strictly theoretical point of view, however, the neoclassical vision of a self-regulating market system did not change (Leijonhufvud 1976) beyond the group of converted Marshallians and later post-Keynesians. That is, the paradigm as a world vision changed only for the narrowest (Marshallian) scientific community, the only one that had full access to Keynes' language and method. In contrast, the neoclassical synthesis absorbed those remarks of Keynes that were less dangerous and went back to the world of equilibrium and rational agents.

Did it change economists' habits and procedures? After the Keynesian revolution economists divided economics between micro and macro. A problem classified in the latter segment would be dealt with by the tools of the ISLM and national accounts. Even now, the existence of a field called macroeconomics is accepted. Despite this, Keynesian economists used risk more than uncertainty, and based their arguments on price rigidity (i.e. amarket imperfection) in their everyday work. This was the result of the encounter of the Keynesian ideas with the formalist tendencies that influenced the theory-inclined members of the profession considered in the second definition of scientific community. The price of this encounter was the loss of concepts associated with words incompatible with the new formalist language.

Did it change the accepted problem-tools? The emphasis on employment and what we now call 'macroeconomic' problems suggests that there was a change here too. The ISLM model became the standard toolkit. But the tools were only those designed by Keynes as a first approximation (e.g. assuming given expectations, rigid prices and stable curves), and the problem disappeared with World War II.[8] The change in the toolkit affected the scientific community in the broadest sense and matched the change of language with which policy-makers were experimenting.

Understanding the Keynesian revolution is a difficult process that requires a detailed and careful historical study. The original sense of paradigm seems appropriate for this task when complemented with Kuhn's later reformulation. The success of the revolution was limited to the scientific community in its

narrowest sense and faded when the circle extended and met with other paradigms. The discontinuities associated with a change in paradigm were different, depending on the distance from the centre.[9]

Thus it seems that this was really a scientific revolution, though all the terms have to be adapted to the features of economics. The revolution's difficulties in fully permeating the scientific community in its broadest sense do not mean that it failed. Post-Keynesian, neo-Ricardian and certain Marxist approaches are developing different (not always coherent) aspects of the Keynesian paradigm towards a creation of an alternative to neoclassical economics. What we have seen is only the first stage of a process that has not yet finished. The Copernican revolution required almost two centuries to impose itself; why, in economics, should we expect a complete change of paradigm in less than fifty years?

ACKNOWLEDGEMENTS

I am grateful to Colciencias and Banco de la República for financial support, and to Geoff Harcourt, Gay Meeks and David Dequech for their comments on a previous draft.

NOTES

1 The debate between philosophers of science will not be studied here. It was mainly directed towards the descriptive/normative character of Kuhn's work, the'irrationality' of scientific work and the definition of a paradigm. (Lakatos and Musgrave 1970; Lakatos 1978).
2 A Kuhnian explanation of the Keynesian revolution was also proposed in a short article by Stanfield (1974).
3 This does not deny the importance of Kahn, but merely stresses the way in which he exerted his influence on Keynes. Kahn, of course, had his own particular ideas that did not come from Marshall, such as his rejection of the quantity theory (Skidelsky 1992:288; Harcourt 1994). Kahn also helped Keynes to avoid the 'foolish things [that] one can temporarily believe if one thinks too long alone' (Keynes 1936:xxiii).
4 The role of these tendencies has been stressed by Leijonhufvud (1976, 1997).
5 For other countries, see Colander and Coats 1989.
6 This was possible because the implicit author was a scholar, not a scientist (McCloskey 1991).
7 For a different view of the cosmology of Keynes' revolution, see Verdon 1996: ch. 5.
8 Other tools have survived. Keynes' inclusion of expectations into the determination of equilibrium became, in the hands of Hicks, the expectations method, which has transformed the notion of equilibrium from long-period positions to temporary equilibrium. The consumption function has been studied as a case of a change of problem-tool paradigm by Argyrous (1992) and Dow (1994).

9 Despite the formal similarity, the centre of the paradigm is not the core of a Lakatosian research programme.

REFERENCES

Adarkar, B.P. (1935) *The Theory of Monetary Policy,* London: P.S.King & Son.

Argyrous, G. (1992) 'Kuhn's paradigms and neoclassical economies', *Economics and Philosophy,* 8 (2):231 – 48.

Barber, W.J. (1996) *Designs within Disorder: Franklin D.Roosevelt, the Economists, and the Shaping of American Economic Policy, 1933 – 45,* Cambridge: Cambridge University Press.

Barnes, B. (1982) *T.S.Kuhn and the Social Sciences,* London: Macmillan.

Blaug, M. (1980) *The Methodology of Economics,* Cambridge: Cambridge University Press, 1992.

Booth, A. (1983) 'The "Keynesian revolution" in economic policy making', *Economic History Review,* 36 (1):103 – 23.

Bradford, W. and Harcourt, G.C. (1997) 'Units and definitions', in G.C.Harcourt and P.Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 1, London and New York: Routledge, 107 – 31.

Brofenbrenner, M. (1971) 'The "structure of revolutions" in economic thought', *History of Political Economy,* 3 (1):136 – 51.

Burghardt, G. (1975) 'A quantitative study of the breakdown and resurrection of the monetary tradition in economies', *History of Political Economy,* 7 (3):360 – 71.

Coates, J. (1997) 'Keynes, vague concepts and fuzzy logic', in G.C.Harcourt and P. Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 2, London and New York: Routledge, 244 – 57.

Coats, A.W. (1964) 'The American Economic Association, 1904 – 29', *American Economic Review,* 54 (4):261 – 85.

——(1967) 'Sociological aspects of British economic thought (ca. 1880 – 1930)', *Journal of Political Economy,* 75 (5):706 – 29.

——(1969) 'Is there a "structure of scientific revolutions" in economics?', *Kyklos,* 22: 289 – 94.

——(1985) 'The American Economic Association and the economics profession', *Journal of Economic Literature,* 23 (4):1697 – 727.

——(1991) 'The learned journal in the development of economics and the economics profession: the British case', *Economic Notes,* 20 (1):89 – 116.

Coats, A.W. and Coats, S.E. (1973) 'The changing social composition of the Royal Economic Society 1890 – 1960 and the professionalisation of British economies', *British Journal of Sociology,* 24:165 – 87.

Colander, D.C. and Coats, A.W. (1989) *The Spread of Economic Ideas; Keynesianism among Different Countries,* Cambridge: Cambridge University Press.

Colander, D.C. and Landreth, H. (1996) *The Coming of Keynesianism to America: Conversations with the Founders of Keynesian Economics,* Cheltenham and Brookfield: Edward Elgar.

Currie, L. (1938) 'Some theoretical implications of J.M.Keynes' *General Theory',* in *The Economic Doctrines of John Maynard Keynes,* New York: National Industrial Conference Board.

Darity, W. Jr and Young, W. (1995) 'IS-LM: an inquest', *History of Political Economy,* 27 (I):1 – 41.

Davidson P. (1987) 'User cost', in J.Eatwell, M.Milgate and P.Newman (eds), The *New Palgrave: A Dictionary of Economics,* vol. 4, London: Macmillan, 766 – 7.

Davis, J.R. (1968) 'Chicago economists, deficit budgets, and the early 1930s', *American Economic Review,* 58 (3):476 – 82.

Deane, P. (1984) 'Paradigms in economies', mimeo, Cambridge University.

Dillard, D. (1978) 'Revolutions in economic theory', *Southern Economic Journal,* 44 (4): 705 – 24.

Dow, S.C. (1994) 'Kuhn's paradigms and neoclassical economics: a comment', *Economics and Philosophy,* 10(1):119 – 22.

Fouraker, L.A. (1958) 'The Cambridge didactic style', *Journal of Political Economy,* 66 (1):65 – 73.

Hall, P.A. (ed.) (1989) *The Political Power of Economic Ideas; Keynesianism Across Nations,* Princeton NJ: Princeton University Press.

Harcourt, G.C. (1994) 'Kahn and Keynes and the making of *The General Theory', Cambridge Journal of Economics,* 18(1):11 – 24.

Harcourt, G.C. and P.Riach (eds.) (1997) *A Second Edition of The General Theory,* 2 vols, London and New York: Routledge.

Harris, S.E. (ed.) (1947) *The New Economics: Keynes' Influence on Theory and Public Policy,* New York: Knopf.

Harrod, R. (1974) *Foundations of Inductive Logic,* London: Macmillan.

Humphrey, T.M. (1973) 'Empirical tests of the quantity theory of money in the United States, 1900 – 30', *History of Political Economy,* 5 (2):285 – 316.

Karsten, S.G. (1973) 'Dialectics and the evolution of economic thought', *History of Political Economy,* 5 (2):399 – 419.

Katouzian, H. (1980) *Ideology and Method in Economics,* London: Macmillan.

Keynes, J.M. (1930) *A Treatise on Money,* London: Macmillan.

——(1936) *The General Theory,* London: Macmillan.

——(1973a) *The General Theory and After: Part I, Preparation,* vol. 13 of *The Collected Writings of John Maynard Keynes,* London: Macmillan.

——(1973b) *The General Theory and After: Part I, Defence and Development,* vol. 14 of *The Collected Writings of John Maynard Keynes,* London: Macmillan.

Keyserling, L.H. (1987) 'Pump priming', in J.Eatwell, M.Milgate and P.Newman (eds.) *The New Palgrave: A Dictionary of Economics,* vol. 3, London: Macmillan, 1074 – 5.

King, J.E. (1994) 'A conversation with Paul Davidson', *Review of Political Economy,* 6 (3):357 – 79.

Klein, L. (1968) *The Keynesian Revolution,* London: Macmillan.

Kregel, J. (1997) 'The theory of value, expectations and Chapter 17 of *The General Theory',* in G.C.Harcourt and P.Riach (eds) (1997) *A Second Edition of The General Theory,* vol. 1, London and New York: Routledge, 261 – 82.

Kuhn, T. (1962) *The Structure of Scientific Revolutions,* Chicago: University of Chicago Press.

——(1970) 'Reflections on my critics', in I.Lakatos and A.Musgrave (eds) *Criticism and the Growth of Knowledge,* Cambridge: Cambridge University Press.

——(1974) 'Second thoughts on paradigms', in *The Essential Tension,* Chicago:The University of Chicago Press , 1977.

Kunin, L. and Weaver, F.S. (1971) 'The structure of scientific revolutions in economies', *History of Political Economy,* 3 (2):391 – 7.

Lakatos I. (1978) *The Methodology of Scientific Research Programmes,* Cambridge: Cambridge University Press.

Lakatos, I. and Musgrave, A. (1970) *Criticism and the Growth of Knowledge,* Cambridge: Cambridge University Press.

Lawlor, M.S. (1994) 'The own-rates framework as an interpretation of the *General Theory,* a suggestion for complicating the Keynesian theory of money', in J.B. Davis (ed.) *The State of Interpretation of Keynes,* Dordrecht: Kluwer, 39 – 90.

Lawson, T. (1997) *Economics and Reality,* London and New York: Routledge.

Leijonhufvud, A. (1976) 'Schools, "revolutions" and research programmes in economic theory', in S.Latsis (ed.) *Method and Appraisal in Economics,* Cambridge: Cambridge University Press, 65 – 108.

——(1997) 'Mr Keynes and the moderns', *Discussion Paper* no. 4, Università degli Studi di Trento, Dipartimento di Economia.

Leontief, W. (1937) 'Implicit theorizing: a methodological criticism of the neo-Cambridge school', *Quarterly Journal of Economics,* 51 (2):337 – 51.

McCloskey, D.N. (1991) 'Mere style in economics journals, 1920 to the present', *Economic Notes,* 20 (1):135 – 58.

McCracken, H.L. (1933) *Value Theory and Business Cycles,* 2nd edn, New York: Falcon Press.

Malcolm, N. (1952) 'Moore and ordinary language', in *The Philosophy of G.E. Moore,* New York: Tudor, 343 – 68.

Masterman, M. (1970) 'The nature of a paradigm', in I.Lakatos and A.Musgrave (1970) *Criticism and the Growth of Knowledge,* Cambridge: Cambridge University Press, 59 – 89.

Mehta, G. (1977) *The Structure of the Keynesian Revolution,* London: Martin Robertson, 1979.

Moggridge, D.E. (1986) 'Keynes in historical perspective', *Eastern Economic Journal,* 12 (4):357 – 69.

——(1995) 'The diffusion of the Keynesian revolution: the young and the graduate schools', *History of Political Economy,* 27 (0) supplement: 223 – 41.

Patinkin, D. (1976) *Keynes' Monetary Thought: A Study of its Development,* Durham NC: Duke University Press.

——(1981) 'Keynes and Chicago', in *Essays in and on the Chicago Tradition,* Durham NC: Duke University Press, 289 – 308.

Patinkin, D. and Leith, J.C. (eds) (1978) *Keynes, Cambridge and The General Theory,* Toronto and Buffalo NY: University of Toronto Press.

Pearce, K.A. and Hoover, K.D. (1995) 'After the revolution: Paul Samuelson and the textbook Keynesian model', *History of Political Economy,* 27 (0) supplement: 189 – 216.

Reddaway, W.B. (1936) 'The General Theory of Employment, Interest and Money', *Economic Record,* 12 (1):28 – 36.

Robinson, J.V. (1933) 'The theory of money and the analysis of output', *Review of Economic Studies,* 1 (1):22 – 6.

Rymes, T.K. (1989) *Keynes' Lectures 1932 – 5: Notes of a Representative Student,* London: Macmillan.

Samuelson, P.A. (1947) *Economics: An Introductory Analysis,* New York: McGraw-Hill.

Sandilands, R. (1990) *The Life and Political Economy of Lauchlin Currie: New Dealer, Presidential Advisor, and Development Economist,* Durham NC: Duke University Press.

Scitovsky, T. (1984) 'Lerner's contribution to economies', *Journal of Economic Literature,* 22 (4):1547 – 71.

Skidelsky, R. (1992) *John Maynard Keynes: The Economist as Saviour 1920 – 37,* London: Macmillan.

Stanfield, R. (1974) 'Kuhnian scientific revolutions and the Keynesian revolution', *Journal of Economic Issues,* 8 (1):97 – 109.

Stein, H. (1969) *The Fiscal Revolution in America,* Chicago: University of Chicago Press.

Stigler, J.C. (1969) 'Has economics a useful past?' *History of Political Economy,* 1 (2): 217 – 30.

Sweezy, A. (1972) 'The Keynesians and government policy, 1933 – 9', *American Economic Review,* 62 (2):116 – 24.

Tobin, J. (1956) 'Liquidity preference as behaviour towards risk', *Review of Economic Studies,* 25 (2):65 – 86.

Turner, M.S. (1989) *Joan Robinson and the Americans,* Armonk NY: Sharpe.

Verdon, M. (1996) *Keynes and the 'Classics': A Study in Language, Epistemology and Mistaken Identities,* London and New York: Routledge.

Tarshis, L. (1947) *The Elements of Economics: An Introduction to the Theory of Price and Employment,* Cambridge MA: Hough ton Mifflin.

Young, W. (1987) *Interpreting Mr Keynes; The IS-LM Enigma,* Cambridge: Polity Press.

13
SITUATIONAL ANALYSIS AND AGENT RATIONALITY
Shackle *contra* Popper

Allen Oakley

INTRODUCTION

Situational analysis (SA) in combination with the rationality principle (RP) in the metatheory of the human sciences is most often associated with Karl Popper (1972, 1976, 1983) and with the critique of their application in mainstream microeconomics by Spiro Latsis (1972). The metatheoretical principle involved is deceptively self-evident: in making decisions about economic actions, agents apply reason to their circumstances in pursuit of their objectives. Nonetheless, the meaning and significance in general, and in economics in particular, of this principle remains the subject of controversy (Koertge 1975, 1979; Latsis 1983; Hands 1991; Nadeau 1993; Lagueux 1993).

The controversy has largely focused on the RP aspect of the combination. The complications and ambiguities raised by the idea of rationality are graphically demonstrated by Popper's inconclusive attempt to wrestle with it (especially in Popper 1983). He pursued a range of circumlocutions in order to try to preserve the *empirical* relevance of the RP while granting that it is but an instrument of logic. He did so in the face of his own conclusion that it can be but a 'minimum principle' because it 'assumes no more than the adequacy of our actions to our problem situations as we see them' (1983: 365, cf. Lagueux 1993). The key ideas here are those of 'adequacy' and 'as we see them', for both imply interpretations and judgements of a subjectivist nature that continually haunted Popper's sometimes objectivist pursuit. One well defended interpretation of the RP has thus rendered it as subjective and unfalsifiable in Popper's terms (Hands 1991; cf. Nadeau 1993). It is argued from this perspective that there is no action that an agent could be observed to take that could refute a claim that it was rational in the sense of being appropriate and adequate to the situation as the agent interpreted it and inaccordance with this agent's pre-existing capacities. If this depiction of the agents' predicament is accepted as arguably an ontologically realistic one, debates about the RP *per se* have little point.

By contrast with this emphasis on the RP, Popper and his critics have been much less concerned to reflect on the nature of the circumstances which agents are

presumed to 'know' and to which they are presumed to apply their reason in reaching decisions to act. Either the circumstances are taken to be axiomatic foundations for the logic that follows, as in mainstream microeconomics: 'It is part and parcel of the neoclassical programme to specify situations which uniquely determine behaviour' (Latsis 1972:211). Or the issue is merely noted in passing as a difficult one (e.g. Koertge 1975: 446). This is a serious shortcoming, for as George Shackle in particular has argued so cogently, the meaning of rationality, as well as the very capacity of agents to be rational, primarily depend upon the character of the circumstances with which they are confronted (Shackle 1958, 1972, 1979). His critical penetration into the deeper recesses of the ontology of human reason and circumstances, and its implications for how agents actually deal with their economic predicaments, is unmatched in the extant literature.[1]

Shackle undertook an extended critical deconstruction of rationality, as it is dominant in mainstream economic theory (1972). This revealed, in particular for our purposes, why it is that in order to understand the reason that agents apply to their circumstances, we need first to understand the true nature of those circumstances. This latter nature is, as we are to see, by no means a simple one. There is much in the circumstances confronted by economic agents which, Shackle argued, defies the application of reason as a process that can be represented as pure logic. But as he realised, the dilemma remains that reason is the agents' crucial cognitive capacity for deliberative decision-making. Understanding economic actions through the SA-RP approach, therefore, requires that we account for the fact that agents somehow adapt their reason to the intractabilities of their circumstances.

In the next section, I give a brief summary of the extant notions of SA and its RP that have followed from Popper's inconclusive seminal arguments and Latsis' critique. My intention in the following section is, then, through Shackle, to elicit the nature of the difficulties inherent in the SA-RP metatheory as these especially stem from the problematic of agents' circumstances. Shackle's ideas on the actual role of reason in the face of this realistic appraisal of agents' circumstances is the subject of the fourth section. The present context allows me only to indicate how future research may productively pursue Shackle's insights by drawing on some of Popper's further ideas about human science and on the agency and structure literature of social theory. This is done in the final section.

POPPERIAN SITUATIONAL ANALYSIS

The thrust of the SA approach to understanding agents' decision-making and action is to depict these as sequential stages in a *rational* response to a problem and its circumstances that are deemed to warrant attention. The response is more extensively comprised of a reasoned appraisal of relevant situational signals and conditions by the agent, a deliberation procedure that facilitates the reasoned

choice of a resolution of the problem confronted, a decision to act (or not to do so) and the implementation of the action (Latsis 1972; Koertge 1975).

The external inputs to reasoned deliberations are the 'known' circumstances confronted by agents, appropriately appraised, classified and selected for attention in the light of the problem to be dealt with. The constitution of what is argued to be 'known' by agents is crucial here. It has two dimensions that should not be conflated. These are most aptly thought of as what agents can know, and what they are actually presumed to know. The former is an epistemological matter that is closely tied to the ontology of the agents' circumstances, while the latter is an operational matter which is, ostensibly, itself the subject of deliberations about knowledge acquisition. An understanding of deliberations generally must be founded on an acceptance of the fact that agents can only apply reason to their circumstances as they, themselves, perceive and understand them. This qualification requires the presumption that agents bring some minimum of accumulated knowledge, reason and other cognitive skills to the selective appraisal of their situations. Whatever the resulting content of what agents believe they know, such appraisal is antecedent to deliberations and decisions about the problem confronted (cf. Koertge 1975:443ff). Such claimed knowledge and understanding of circumstances is the substance to which the processes of reason are applied in search of a choice which will best resolve the perceived problem.

If rationality is to have any operationally meaningful status from the agents' perspective, it must involve them in confronting and dealing with these realities concerning the nature and use of knowledge of their circumstances.[2] The notion that rationality can mean that agents could, even potentially, deal perfectly (optimally) with circumstances *as they in fact are* is clearly false as an objective or empirical matter (Nadeau 1993:459). The reasons for this are, first of all, practical, in that no amount of effort or resource expenditure could ever discover all the relevant facts of circumstances. The idea is also epistemologically dubious, in that there are few, if any 'facts' as such because circumstances are as interpreted by individuals. Popper tried to side-step the difficulty at one point: 'having constructed our model, our situation, we assume no more than that the actors act within the terms of the model, or that they "work out" what was *implicit* in the situation' (Popper 1983:359, original emphasis). However he was also prepared to refer to it directly:

> In order to understand…[agents'] (inadequate) actions, we have…to reconstruct a wider view of the situation than their own. This must be done in such a way that we can see how and why the situation as they saw it… led them to act as they did; that is to say, adequately for their inadequate view of the situational structure.
>
> (Popper 1983:363)

However, as Shackle has detailed for us, a deeper epistemological difficulty than this is involved in understanding agents' circumstances. For when the problem confronted by agents requires ex-ante decisions, as so many significant economic decisions do, the very idea of knowable facts becomes a dubious one. Decisions about the future can be reasoned out on the basis of nothing but creative imagination and expectation informed by knowledge and experience. The deliberated choices made on such bases come from an open-ended set of alternative scenarios that cannot be delimited except by reference back to the agent's reasoning about feasibilities as they believe them to be under the circumstances as they understand them. This is the highly subjective image of operational rationality to which Popper drew our attention in the above passage, but to grasp its full import, we need to turn to the contributions of Shackle.

SHACKLE *CONTRA* POPPER[3]

For Shackle, as for Popper, the idea of understanding the human dimension of economics on the basis of assumed agent rationality within a set of specified circumstances had a certain common-sense attractiveness. 'Reason has attracted man since the Greeks, because it was something that a man could find within himself (Shackle 1972:80). Thus: 'Reason is sure, safe, even in a sense simple. The economic analyst has opted for Reason. His guide is a single principle. He assumes that men pursue their interest by applying reason to their circumstances' (1972:xii). It was no surprise to Shackle that theorists endeavoured to hold onto the approach, in spite of charges of unrealism, and to pursue it in a form that brought logical rigour and predictive reliability to the representation of economic phenomena. 'An axiomatics, to attain beauty, cogency and usefulness, needs a certain simplicity. Thus the first policy-decision of the constructors of [economic] theory, [was] to assume that conduct was governed by reason and could be understood by it' (1972:444).

But Shackle had profound misgivings about the intellectual *bona fides* of the SA-RP combination in its logical format, arguing that it necessarilyobscures the very substance that makes economics a human science. 'Economics has gravely and greatly misled itself by a tacit belief that *rational self-interest* is as simple a basis of prediction as the laws of physical motion' (1972:37, original emphasis). The analytical achievements were, he thought, bought at too high a price in terms of the loss of ontological integrity in interpreting and understanding the economic decision-making and conduct of human agents. The cost resides in what, by its nature…[economics] is obliged to neglect or even implicitly to declare unimportant…. Having opted for the supremacy of reason, it rejects what conflicts with reason' (1972:443ff). In particular, economics has neglected to pursue realistically the nature of the circumstances confronted by economic agents and the consequent limitations on what it can mean to be rational.

The methodologically driven nature of this neglect is evident in the treatment of human reason as pure logic and the confinement of the circumstances to which reason is applied to premises that are axiomatically expressed, complete in their scope and assumed to be fully known.

> The theoretician assumes that men pursue their interests by applying reason to their circumstances. He assumes that they are both rational and fully informed. Given a complete statement of all a man's relevant circumstances, the theoretician can, by this method and this basic assumption, predict what a man will do.
>
> (Shackle 1972:4, cf. 239)

The supposition that agents pursue self-interest by applying reason to their circumstances in such theory is the equivalent of arguing that *their circumstances directly govern their actions* (cf. 1972:86). 'Reason…is the fitting together of ideas. The strong temptation of a theory of determinate conduct is to suppose, without argument, that the ideas which reason fits together are in some way *given*' (1979:57, original emphasis). And that the circumstances are pre-given in such a way as already to embody the outcome, just as the pieces of a jigsaw are cut from a picture that exists beforehand. 'Conduct, this view assumes, is the fitting together of a jigsaw puzzle. The pattern in which this can be done is unique, determined by the shapes of the pieces. Not only the need for the pieces to fit together, namely the operation of reason, but also, in the determinist view, the pieces themselves are given' (1979:57). All this is potentially misleading in that it is apt to confuse the confines of logical argument and pre-established premises with the actual determination of the conduct it purports to represent.

Nonetheless, it remains the case, Shackle reminded us, that 'Imagination and Reason are the two faculties that make us human', for 'every hour and every moment provides evidence that we pursue reason with unflagging ardour and indefeasible instinct and intuition' (1979:xii, 56). It is

> human nature [that] impels men to apply reason to their affairs, the human situation turns those affairs into problems of policy and action which overwhelm the fragmentary, doubtful, disordered and often seemingly self-contradictory evidence and suggestion which the general scene supplies for their reason to work on.
>
> (Shackle 1972:272)

As Shackle observed, however, in this respect

> the theoretician has seldom asked himself how this supposedly complete knowledge of circumstance could be attained by the individual. Which of the economist's theories have looked in the face of the genuine

> predicament of man, the logical non-existence of such complete knowledge?
>
> (1972:4)

And, he continued,

> an important and essential part of…[the] governance [of conduct] is exerted not directly but through men's knowledge and thoughts…. The study of the pure application of reason drove out and logically excluded the question of the *possibility, nature* and *source* of knowledge.
>
> (1972:86, emphasis added)

For Shackle, then, understanding the substance and import of the SA-RP metatheory involved three facets: the existential nature of agents themselves, defining the nature of their circumstances and establishing the meaning of the agents' practical application of reason to their circumstances. He showed that none of these facets can legitimately be confined to the narrow and objectified scope and meaning apparent in logically deterministic SA-RP based models. Each facet requires careful elaboration if ontologically grounded understandings of economic phenomena are to be developed.

The existential image of agents that provided the backdrop for Shackle's inquiries depicted them as inherently volitional and contingent, but as shaped and directed in their conduct by the problems that they address and the external situations within and through which they act. His focus was on a vision of economics in which what such agents think, feel and do in order to improve their well-being to the maximum extent they believe possible, given their situations and resource constraints, were given primacy (1972: xviii, 246, 256).

It is apparent that human agents express emotions and feelings, and that they make judgements and follow prejudices and motives, as well as being creative, and originating, when they formulate and pursue purposes andgoals. As Shackle asked: 'What is action but the response to feelings? What action would there be if there were no desires, no consciousness of dissatisfaction, no longing for a "good state of mind"? What is *motive* except emotion?' (1972:135, original emphasis). Agents bring these subjective characteristics to their deliberations. 'We choose, and take, action in pursuit of an end. What is an end if not something upon which our desire concentrates our thought and effort? Again, what is desire if not emotion?' (1972:135). So, he argued, 'economics is the analysis of that conduct, and its results, which originate from what people look forward to; conceive, upon fragmentary evidence, as possibly attainable…conduct springing from imagination and hope' (1972:66). He referred to this pursuit as the ex-ante 'maximum attainment of *a good state of mind,* a good state, that is, of *imagination,* of the conception of states to be attained which cast the warmest glow of anticipation on the present' (1972:xviif, original emphasis, cf. 1979:45).

In this respect, 'choice erects a structure of intentions, any abandonment of which will be hurtful to the chooser in some degree' (1979:15).

Shackle accepted that operationally, agents deliberate about their actions by applying reason to their circumstances in order to pursue their objectives. In doing so, he implicitly posited a dual problematic concerning those circumstances upon which understanding of deliberations and decisions to act must depend. First, what can be said about the nature of circumstances *per se* as objects of potential knowledge? And second, what knowledge can agents realistically be expected actually to obtain about such circumstances? Clearly the second argument depends very much upon what is established about the first. But this cannot be the full story because of the mediation of two additional factors. One is the amount of resources agents devote to obtaining knowledge. The other, less tractable factor is the variable innate and learned cognitive and creative capacities that agents bring to the interpretation process which gives knowledge its practical relevance. Moreover, the processing of knowledge that comprises deliberation as the manifestation of practical reason depends on these capacities, too. This complicates the knowledge issue even further. All of this adds up to a highly subjectivist vision of the SA-RP approach. But, whatever its awkward implications for theory construction, its enduring merit is that it confronts squarely the agent ontology which accounts for the generation of economic phenomena.

To some limited extent, the knowledge of circumstances with which agents work can be improved by the expenditure of resources, both human and pecuniary. This applies to those dimensions of circumstances which have a material or discursively recorded existential form to which agents can gain access. Deliberation demands that agents apply accumulated knowledge, reason, other cognitive capacities and ingenuity to their circumstances, as they perceive them. This process is thus bounded by the capacities and the resources which are apparent to the chooser in themself and are within their reach in the period of deliberation (1979:31). 'A *course of action by thechooser* involves his dedication of particular resources, available to him in limited quantities, to specific kinds of use…. Action involves means, means are limited in quantity, capacity, power and reach, for any chooser' (1979: 37, original emphasis).

It was apparent to Shackle that the uncertainties and ambiguities concerning knowledge confronted by agents cannot be overcome by resource use alone. He asked pertinently: 'Where, in real life, are we justified in assuming that we possess "enough" information? What would be the test of "enough"? How many years would have to elapse [after a decision and action] before that test could be conclusively applied?' (1972:184). Further intricate difficulties are involved, too: The worth of new knowledge cannot begin to be assessed until we have it. By then it is too late to decide how much to spend on breaching the walls to encourage its arrival' (1972:272f). The situation is really even more intractable than this, for agents cannot assess the worth of information until *they actually*

use it for their deliberations and test its effects as results of actions. But this problematic confronts the general impossibility of any assessment of how things could have been otherwise, if more or less had been expended, or if what was expended had been directed differently. To this could be added the observation that it would all be to no avail ex post, anyway, for it is simply too late to change anything, and next time a decision of the same type is made, all the information inputs will be different: we cannot 'learn' much in such circumstances (cf. 1972: 86). Most significantly, economic actions often constitute for agents what Shackle called '*crucial* experiments' or '*self-destroying experiments,* in which the performance of the action, the making of the experiment irreversibly destroys the conditions which are its essence' (1979: 135, original emphasis).

What all this meant for Shackle was that the reasoning and deliberations of agents could not be represented by a closed logical argument. As he put it,' "Rationality" is an empty and idle term until the data available to the individual are specified', for 'to act by reason, a man must be fully informed of his circumstances so far as they bear on the outcome of his actions' (1972:37, 91). The problem, as we have seen, is that these data are so ill defined. So, concluded Shackle:

> Reason unfolds the meaning of the premises. It can do no more. But if one of the premises should be: There are things unthought of, that time in its operation will suggest, what can reason tell us about that entirety, which contains both visible premises and the unknown things?
>
> (1972:25)

Moreover, as a practical matter, 'men act, not upon calculation but upon suggestion, building upon foundations which they conjure from one ofmany or infinitely many possible interpretations of what they see' (1979:76). Thus, 'to be human is to be denied the necessary condition of rationality, complete relevant knowledge. So far is knowledge from ever being complete, that consciousness consists in the gaining of knowledge' (1972:446).

There exists a crucial problem about the nature of circumstances in themselves which, for Shackle, pointed beyond the issues to which I have just referred. This comes about most obviously because so much about which agents must deliberate and decide relates to the temporal dimension and time-experience of their existence. In Shackle's view, it is time that constitutes for the economist 'the very nature and being of his subject matter' (1972:286). It follows that 'in order to reflect the human predicament', the analyst must 'consider time as the fact above all facts, conditioning every thought, act and meaning' (1972:255). With their orientation towards future pecuniary and/or non-pecuniary returns, economic problems and decisions must stem from deliberations in which real time is the structuring medium. In all such cases, the actions chosen will deliver returns to agents only in the future, so time becomes of the essence in the

deliberations leading up to choice.[4] As a consequence of this non-excludable temporal dimension in agents' deliberations, the *knowledge of and information about* existing circumstances is not only always 'incomplete' because of resource limitations and the other conundrums referred to above. It is also insufficient and non-completable because of the inherent open-endedness introduced by contemplation of what cannot be known about the future. That is, agents confront a state of *unknowledge* of all future circumstances that may be relevant to the outcome of current actions when deliberating about the choices that will result in those actions. Future knowledge is not merely uncertain; it is rather represented as unknowledge by Shackle because there are no means of potential mitigation that are suggested in the semantics of the notion of *un*certainty.

Shackle referred to the continuity of time *per se* and implied that the links are provided by agents: 'Time-present, time-to-come, are not divided by a boundary but are continuously involved one in the other and born of each other, the very being of time-present consists in its imaginative engendering of merging phases which carry it into its successor' (1979:47). In operational terms, he noted that as agents, 'we are prisoners of the present who must choose in the present on the basis of our present knowledge, judgements and assessments' (1979:89). But there is the additional consideration here that 'decision means literally a cut… between "past and future"' (1979:56) in the sense that any determining link between a potentially knowable past and an unknowable future is blocked by what may be called an epistemological wall. This barrier enters the agent's perspective to distinguish the nature of thought about the past and forethought about the future. The epistemic quality of the two facets of thought are very different, as Shackle well understood in his references to agents' cognitive responses inthe face of unknowledge about the future. The confrontation with time-experience and unknowledge of circumstances renders their application of reason ineffective at the critical point in their deliberations. For Shackle, 'time is a denial of the omnipotence of reason…[for it] divides the entirety of things into that part about which we can reason, and that part about which we cannot' (1972: 27). More specifically, '*time* and *complete knowledge* are utterly incompatible. Time is what brings *new* knowledge: how then could the old knowledge ("old" at any given moment) be complete?' (1972: 151, original emphasis).

Agents cope with the pervasive, future-oriented aspect of their circumstances and deliberations by forming expectations. And expectations can be no more than imagination on their part, conjured up out of mental character, knowledge, experience, creativity and delimited cognitive capacity. Although, as Shackle asserted, 'expectation is a structure cantilevered out from the present over a void' (1972:179), it is a void that demands be filled by choice of action: 'Time-to-come imperiously demands to be filled, it provides the practical business of imagination' (1979:8). Because imagination precedes choice, 'deliberative conduct, *choice,* the prime economic act, depend for their possibility…upon the conceptual power of the mind. Choice is necessarily amongst thoughts, amongst

things imagined' (1972: 130, original emphasis). Shackle's critical point here was that 'expectation is origination in process. The nature and number of its potential products are, at any moment, unknowable' (1972:367). The fact that all agents must accept, then, is that the possible range of alternative imagined scenarios, the 'skein of rival hypotheses' for economic decisions and their outcomes that agents are able to create, must 'remain incomplete and incompletable' (1983: 116, cf. 1979:17, 56f, 80ff). Thus, he concluded, 'expectation and determinacy are incompatible and mutually exclusive'. The rational ideal, such as that expressed in the orthodox SA-RP, 'must therefore exclude expectation, and expectation, since it is real and insistently present and accompanies all activity, must destroy the rational ideal' (1972:156).

THE ROLE OF REASON UNDER THE CIRCUMSTANCES

The immediate dilemma much emphasised by Shackle in the above outline of his ideas is that while imaginatively created circumstances and reason as logic do not mix, reason remains the crucial cognitive means that agents must use for ordering and focusing thought. It is in this mediating role of imagination and creativity, demanded by time and unknowledge, between circumstance and action, that Shackle locates the contingency and consequent analytical intractability of human agency.

The only mitigation that may assist formal understanding is that the voidof future time, comprising the period to the horizon affected by present decisions, can only reasonably be filled by what agents believe to be choices that result in actions and outcomes that are possible. 'Imagined filling of time-to-come must, in order to serve its purpose, be deemed possible…. As a tool of the inescapable task of choosing action, imagination must observe constraints' (1979:8). In particular, imagination is contained by the unavoidable need for agents to plan and operate within and through selected, purpose-relevant structures and relations of their inherited and given situations. Situations comprise more or less objective structures and inter-agent relations which *facilitate and contain what is feasible for agents to do*.[5] Any formal understanding of agency can at best, then, rely on understanding the situations and relations confronted by agents and the way such structures shape and regularise their conduct. The result is that such conduct is rendered typical for the circumstances as far as the observer is concerned. It is here that we have come full circle back to the SA-RP approach, in that agents' conduct is argued to be the consequence of their cognitive processing of the imagined and interpreted circumstances they confront. But now the approach is driven by concerns about ontological integrity in understanding reason and circumstances, rather than by methodological imperatives.

In this endeavour to ground the formal representation and understanding of agents' economic decision processes and actions on the direction embedded in volitional situational facilitation and containment, Shackle provided us only with

some hints. Imagined, deemed-possible future scenarios must conform to 'epistemic possibility [which] is *a judgement* that some evolution of affairs in time-to-come is free of fatal obstacles *discernible to the chooser*' (1979:36, emphasis added). For 'the field does not, cannot, offer *existential rivals*. It can only suggest rival evolutions of affairs' (1979:12, original emphasis). Such suggestions from 'the field' originate in the agents' biographical experiences and situations:

> They are the touching of the keys on the keyboard of personal experience, of structures of explanation which the individual has built up for himself, by some resemblance or reminiscent impression reported in the present. They depend upon his own individual history, the detail and exact content of his educative life. Suggestions are evocations of a *personal* response.
>
> (1979:151, original emphasis)

According to Shackle, among imagined scenarios that suggestion can conjure up, the

> adjudged capacity-to-be-realised has three separately indispensable bases. Any such…business of choice is required to conform to Nature and human nature, it must respect the principles of thearchitecture of the field…[and] since 'time-to-come' takes its start from the chooser's present, a history for that time must be an evolution *starting from the situation at that present, as the chooser sees it*.
>
> (1979:13f, emphasis added)

From the agents' perspective, 'at any present, the material means of action are arranged in some pattern as the upshot of past history up to that present', with the result that 'the acts open to the chooser…are infinitely various, but various only within bounds set by the present physical configuration of these means' (1979: 25). At one point, Shackle's consciousness of the situational and experiential containment that applies to operational agents led him to make the following concession with respect to the intractability of ex-ante reason:

> [U]nknowledge is not absolute. The chooser, we suppose has his scheme or geometry of natural principles which set the bounds of *natural possibility*. He has some picture of the particular contingent state to which history has brought 'the present', a set of circumstances which, within the bounds of natural possibility, impose the limits of *general possibility,* the totality of all the skeins of imaginable sequels of the choices of action open to him, for which sequels he can discern no fatal obstacles.
>
> (1979:95, original emphasis, cf. 27ff)

It is, then, from within these bounds that agents can be expected to choose what Shackle called their 'special possibility' skein of feasible sequels: 'the mutually rival imagined sequels which some one choice of action seems to leave not discernably obstructed but which other choices would preclude' (1979:95).

CONCLUSION AND FUTURE DIRECTIONS

The challenge from here on is to pursue more extensively and intensively Shackle's insights with respect to the influence of situational conditions on agents' reckoning of the future scenarios they deem to be possible. This would take inquiries towards developments in social theory that give the substance of the SA-RP metatheory due recognition as an integral part of the general problematic of agency and structure (Giddens 1979: Hindess 1988; Hollis 1977, 1994; Sztompka 1994). There is a sense, too, in which this would bring Shackle's argument back into contact with Popper's cogent analyses concerning the problem of scientific understanding when the complex balance of agents' freedom of action and the control of situational conditions is under study (Popper 1972: ch. 6). This relatively neglectedcontribution by Popper has been given some attention by Latsis (1983), but it has not since been pursued in relation to economics.

Popper and the social theorists have suggested that the means of mitigating the inherent subjectivism of the SA problematic can be ontologically grounded in the agents' own ways of dealing cognitively with their problems and situational conditions. The objective of further metatheoretical inquiry in this direction would be to show that within the exigencies of subjective agency, there exists a balance of situational influences of facilitation and constraint, and of individual contingent freedom. It can then be argued that this immanent form of agent control and direction, of volitional acceptance of situational conditioning of conduct, is the means by which agents themselves avoid the potential nihilism of their own contingency and existentially isolated predicament when making decisions. There will probably always be a contingent remainder in any conduct profile, but the typical is what is to be extracted, and not the individual in the strictest sense. This was evident in Popper's own limited expectations of the human sciences because they 'operate almost always by the method of constructing *typical* situations and conditions'. So, he concluded, SA 'provides us with…rough and ready models…of typical social situations. And my thesis is that only in this way can we explain and understand what happens in society' (Popper 1983:357, 358, original emphasis). Nonetheless, if ontological integrity is to be an objective of representational theory in economics, it is arguably apt to ground theory on an endeavour to understand how agents deliberate about their actions. The intellectual credentials of such an approach are well established elsewhere in the human sciences (cf. Hindess 1988).

ACKNOWLEDGEMENTS

I am grateful to D.Wade Hands and Jeremy Shearmur for their critically penetrating but helpful comments on some of the ideas in this paper included in its earlier guise as Oakley 1993. I remain, nonetheless, solely responsible for their present form.

NOTES

1 My concern here is with what Shackle had to say about the inherent difficulties of understanding the deliberations and actions of economic agents in terms of reason applied to circumstances. I am not concerned with his own efforts formally to represent the processes involved. These have already received much critical scrutiny in the literature (cf. e.g. Ford 1994:70ff).

2 I leave aside considerations of rationality that relate to individuals who by the accepted standards of the time are said to be eccentric or mentally subnormal.For the present purposes that relate only to non-clinical human sciences, such individuals are not significantly relevant.

3 Those familiar with Shackle's vast bibliography (see Ford 1994:490ff) will realise that I can only provide what I believe to be a sketch of his essential ideas on this theme by focusing on two of his most important works. In his *Imagination and the Nature of Choice* (1979), for example, setting out his theory of agents' choice alone occupies more than 150 pages of intensive argument. I must, therefore, leave it to readers to follow up the details of Shackle's ideas that cannot be fully exposited or defended here. However, the one mitigating circumstance is that there is much repetition at the level of essentials in Shackle's many writings on the themes to follow.

4 In the present context, the focus is on short-term decisions taken by agents within a given, inherited set of circumstances and a particular time-horizon of expectation. Longer-term concerns about structural change brought about by collective short-term decisions are not at issue.

5 See note 1 above.

REFERENCES

Ford, J. (1994) *G.L.S.Shackle: The Dissenting Economist's Economist,* Aldershot: Edward Elgar.

Giddens, A. (1979) *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis,* London: Macmillan.

Hands, D.W. (1991) 'Popper, the rationality principle and economic explanation', in G.Shaw (ed.) *Economics, Culture and Education,* Aldershot: Edward Elgar, 108 – 19.

Hindess, B. (1988) *Choice, Rationality, and Social Theory,* London: Unwin Hyman.

Hollis, M. (1977) *Models of Man: Philosophical Thoughts on Social Action,* Cambridge: Cambridge University Press.

——(1994) *The Philosophy of the Social Sciences: An Introduction,* Cambridge: Cambridge University Press.

Koertge, N. (1975) 'Popper's metaphysical research program for the human sciences', *Inquiry,* 18:437 – 62.

——(1979) 'The methodological status of Popper's Rationality Principle', *Theory and Decision,* 10:83 – 95.

Lagueux, M. (1993) 'Popper and the rationality principle', *Philosophy of the Social Sciences,* 23, 4:468 – 80.

Latsis, S. (1972) 'Situational determinism in economies', *British Journal for the Philosophy of Science,* 23:207 – 45.

——(1983) 'The role and status of the rationality principle in the social sciences', in R.S.Cohen and M.W.Wartofsky (eds) *Epistemology, Methodology, and the Social Sciences,* London: D.Reidel, 123 – 51.

Nadeau, R. (1993) 'Confuting Popper on the rationality principle', *Philosophy of the Social Sciences,* 23, 4:446 – 67

Oakley, A. (1993) 'Ontology, realism and economic theory', mimeograph.

Popper, K. (1972) *Objective Knowledge: An Evolutionary Approach,* Oxford: Clarendon Press.

——(1976) 'The logic of the social sciences', in T.W.Adorno, H.Albert, R.Dahrendorf, J.Habermas, M.Pilot and K.R.Popper *The Positivist Dispute in German Sociology,* New York: Harper and Row.

——(1983) 'The rationality principle' [1967], in D.Miller (ed.) *A Pocket Popper,* Oxford: Fontana.

Shackle, G.L.S. (1958) *Time in Economics,* Amsterdam: North Holland.

——(1972) *Epistemics and Economics: A Critique of Economic Doctrines,* Cambridge: Cambridge University Press.

——(1979) *Imagination and the Nature of Choice,* Edinburgh: Edinburgh University Press.

——(1983) 'A student's pilgrimage', *Banca Nazionale del Lavoro Quarterly Review* 36, 145:107 – 16.

Sztompka, P. (ed.) (1994) *Agency and Structure: Reorienting Social Theory,* Yverdon, Switzerland: Gordon and Breach.

14
THE DEBATE ON EXCESS CAPACITY IN THE 1930s

Claudio Sardoni

INTRODUCTION

During the first half of the 1930s, many economists were engaged in a debate on the possibility of having firms or industries in a position of equilibrium characterised by the existence of unused productive capacity. Although this was a microeconomic issue, it may seem strange that a debate on the possibility of equilibria with industries producing less than their potential output was not connected to the contemporary development of Keynes' principle of effective demand and his notion of macroeconomic underemployment equilibria. This may seem even stranger in view of the fact that some of the protagonists of the microeconomic discussion were at the same time deeply involved in the process that led to *The General Theory*. Keynes, however, did not show any serious interest in the debate on excess capacity, and likewise none of the participants in that debate tried to establish a connection with the new theoretical developments in macroeconomics.

The possibility of equilibria with excess capacity was explained by the assumption of imperfect competition. The debate on the relationship between imperfect competition and excess capacity equilibria, both in the short and in the long period, had begun in 1930 with an article by Harrod. However, one year earlier, Kahn had already dealt with the issue in *The Economics of the Short Period* (Kahn 1989), his dissertation for a Fellowship at King's College, Cambridge. But Kahn's work was not published until 1989 and he did not play a public role in the debate of those years.[1]

In 1932, Joan Robinson published an article that seemed to provide a definitive and neat solution to the problem of excess capacity. In imperfect competition, long-period equilibrium *must* necessarily be associated with firms experiencing excess capacity. Her article, however, was soon criticised by Shove and Harrod for being based on too unrealistic assumptions and hence unable to offer a satisfactory explanation of the working of actual markets. In 1933, Robinson published *The Economics of Imperfect Competition* and Chamberlin published his *Theory of MonopolisticCompetition*. In both works, though in a

rather different context, the validity of the 'doctrine of excess capacity' was reaffirmed. Also on this occasion, the doctrine was criticised, by Harrod again and by Kaldor, because it failed to provide a realistic explanation of the behaviour of non-perfectly competitive markets.

The notion of imperfect competition had been introduced as an alternative to the traditional hypothesis of perfect competition because, it was claimed, it represented a more satisfactory description of actual markets. The doctrine of excess capacity, in particular, aimed to give a theoretical explanation of why firms can work below capacity. However, by the mid-1930s, this doctrine had been subjected to serious criticisms concerning its effective ability to provide a satisfactory explanation of the actual working of markets. In the concluding section of the paper, I argue that the unrealistic assumptions on the working of markets and its static approach may have made the 'doctrine of excess capacity' very unpalatable to Keynes, who was trying to develop an economic theory for the 'world in which we live'.

THE NOTION OF EXCESS CAPACITY

Before turning to a more detailed reconstruction of the debate, it is useful to expound clearly what was meant by full and excess capacity at the time of the debate. The concept of full capacity in the prevailing theory of the firm was not defined in physical terms. Kahn, as we shall see presently, was an exception, but others did not follow his approach.

In the short period, for a profit-maximising firm working under decreasing returns, there was no definition of full capacity related to some physical limit to its ability to increase the output. In principle, by increasing the quantities of variable factors, the firm could expand its output to any extent provided that the price was high enough to equal the marginal cost of that level of production. In this context, instead of a physical notion of full capacity, what was generally used was a notion of *optimum capacity,* i.e. the point at which the average total cost reaches its minimum. For any price lower than average total cost, the firm would produce below its full capacity; if the price was lower than the average prime cost, the firm would stop producing and its capacity would be 100 per cent unused. Thus there would exist excess capacity for any level of production smaller than that associated with the minimum average cost. If the firm experiences excess capacity, this implies that it is producing at a decreasing average total cost.

As to the long period, the crucial concept was the notion of *optimum size of the firm*. Given the firm's long-period cost curve (U-shaped as well), its optimum size was reached when it produced its output by using the plant associated with the minimum average cost. In equilibrium the price equals the average cost and the marginal cost; therefore, if the firm produces anamount of output less than

that which could be produced at a minimum average cost, this implies that it experiences decreasing average costs and excess capacity.

KAHN: THE ECONOMICS OF THE SHORT PERIOD

The debate on excess capacity was mainly centred on the long period but there were two exceptions: Kahn, who dealt only with the short period, and Harrod, who dealt with both. In Chapter 5 of his dissertation (Kahn 1989: 45 – 63), Kahn analysed the factors which determine the shape of prime cost curves by questioning the usual assumption of U-shaped curves. He also gave a definition of full capacity in physical terms: a firm produces to full capacity when all its machinery is in use.

Kahn carried out a detailed examination of the way in which firms change their level of output. He listed five possible ways but pointed out that only two methods were those most often chosen:

1 altering the amount of machinery in use;
2 altering the number of working days per week or month.[2]

The choice between the two methods depends on the level of the average prime cost: the method with the lower prime cost is chosen.

Kahn used a diagram similar to the one in Figure 14.1. With the two methods, prime cost is the same at full capacity *(F),* but with method 1 its behaviour to the left of *F* can be depicted by a usual U-shaped curve; with method 2 the prime cost is substantially constant up to *F*. To the right of *F,* where the firm works at full capacity, the prime cost rises vertically. Therefore, the two cost curves associated with the two alternative methods generally meet at two points, *H* and *F*.

The firm chooses the method that, for any level of output, ensures the lower prime cost. Therefore the bold line in the figure above represents the behaviour of the average prime cost. The position of *H,* the intersection point below full capacity, depends on the degree of uniformity of machinery and the incidence of the so-called 'quasi-fixed' costs.[3] If machines are perfectly uniform and 'quasi-fixed' costs are relevant, the cost curve of method 1 would meet the other cost curve only at one point, at full capacity. In this case method 2 would be the more economical for all levels of output and the cost curve would take on the shape of a reverse L.

Kahn held that in many industries firms are characterised by a fairly complete uniformity of machinery, so that an adequate description of the behaviour of their costs was provided by a reverse L-shaped cost curve (Kahn 1989:49 – 59). From this it follows that, in general, average prime cost is independent of the level of output and, therefore, that the marginal primecost is equal to the (constant) average prime cost. Up to the point of full capacity the marginal prime

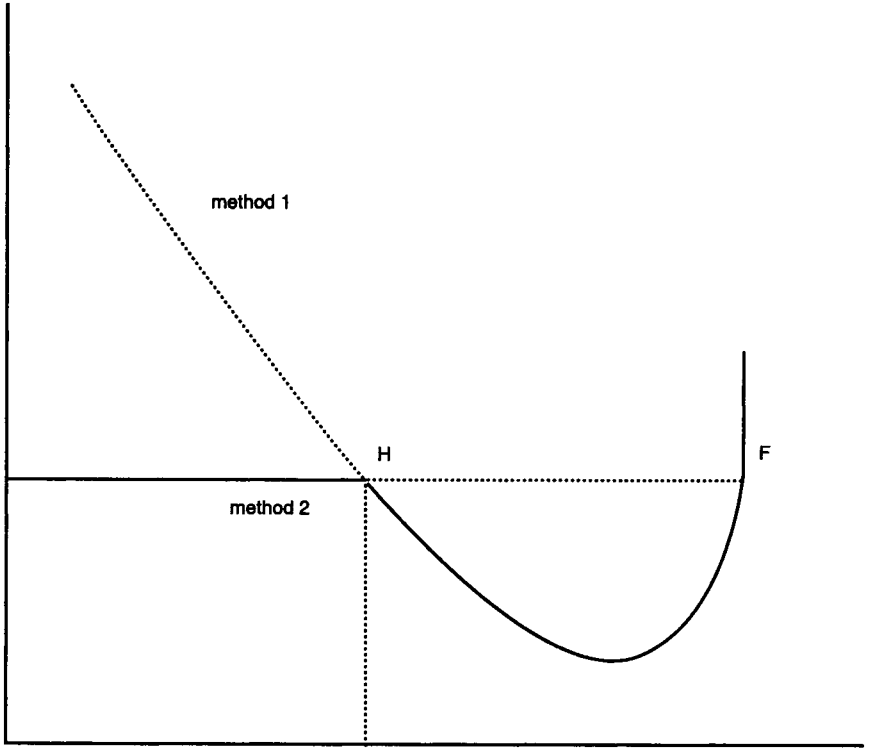

Figure 14.1 Two common methods by which firms change their level of output

cost is constant and then it rises vertically. This had to bring about the *collapse of the assumption of perfect competition.*

A marginal prime cost curve that is perfectly elastic to the point of full capacity is not significant for the determination of the equilibrium level of output. Insofar as the price is above the average prime cost, a firm in perfect competition would produce to capacity; if the price were below the average prime cost, the firm would not produce at all. For Kahn, however, such a conclusion was in contradiction with concrete experience: in reality firms also work by leaving part of their capacity unused. A satisfactory explanation of this fact could be found by abandoning the hypothesis of perfect competition, i.e. the hypothesis that firms face a perfectly elastic demand curve (Kahn 1989:59 – 60). If the demand curve is downward sloping, the firm can earn a profit also by producing less than its maximum output.

HARROD: EXCESS CAPACITY IN THE SHORT AND IN THE LONG PERIOD

In 1930 and 1931, Harrod published two articles (Harrod 1930, 1931) where he dealt with equilibrium in the short and the long period. Harrod first determined equilibrium in perfect competition. The two notions of equilibrium are traditional, but some of Harrod's observations concerning the cost curves are significant in that they apply also to the analysis of imperfect competition.

Harrod held that, both in the short and in the long period, marginal and average cost curves are U-shaped. In the short period, the reward of fixed factors is the Marshallian quasi-rent and, therefore, the supplementary cost cannot be predetermined because the relation between demand price and prime cost passively determines the price payable to fixed factors. In order to give a definite meaning to the supplementary cost and hence to the average total cost, it is then necessary to suppose that fixed factors should receive a 'normal profit', so that the supplementary cost K includes a normal return to fixed resources (Harrod 1930:80). In this way, if the demand curve cuts the marginal prime cost below the average total cost but above the average prime cost, fixed factors are receiving a subnormal reward. In such a situation, the firm is necessarily working below capacity and gets less than the normal profit. The point of optimum capacity is always a point at which marginal costs are rising, and profits are at their normal level (Harrod 1930: 82). Harrod maintained a concept of fixed factors also in the long period, which he regarded as a period of a definite length s that is longer than the short period but shorter than infinity. Over s, short-period supplementary costs become prime but, since factors such as goodwill or part of the fixed equipment survive the period s, their cost remains supplementary also in s. In this way, decreasing returns for 'long-period prime factors' can be explained essentially in the same way as decreasing returns to variable factors in the short period (Harrod 1930:81 – 2).

Harrod then turned to consider a case in which the size of the firm is not small in relation to the size of the whole industry. In this case the firm faces a downward sloping demand curve. He showed that, in this case, there could be equilibrium with decreasing costs and excess capacity while the firm earns a profit (Harrod 1930:84 – 6). This is a fundamental difference between perfect and non-perfect competition.

In 1931, Harrod argued that the phenomenon of firms facing a downward sloping demand curve could not be limited to the case in which the size of the firm is large in relation to the whole industry. In fact, firms, although small in relation to the whole industry, face a downward sloping demand curve also when the product is not perfectly standardised or the market is not completely organised. The downward sloping nature of individual demand curves depends on firms' increasing difficulty in marketing their additional output (Harrod 1931: 566).

Harrod set out to determine equilibrium under these new conditions. The equilibrium is at the point of intersection of the marginal cost curve and the 'increment of aggregate demand curve' (Harrod 1931:570). What Harrod called 'increment of aggregate demand curve' is in fact what Robinson would call 'marginal revenue curve' one year later.[4] But apart from terminology, Harrod's determination of equilibrium is the one that became universally accepted: the firm maximises its profit when marginal cost is equal to marginal revenue. In this equilibrium position, the firm can earn higher than normal profits, as in the case depicted in Figure 14.2. It is a short-period equilibrium, when there is no reason to assume that profits are at their normal level.

He then considered long-period issues. He first argued that an equilibrium characterised by decreasing costs could be regarded as a long-period equilibrium, if there existed considerable marketing and transport costs that more than compensate for decreasing costs of production (Harrod 1931: 97 – 8).[5] Harrod then turned to the question whether conditions of decreasing costs could be regarded as normal for certain industries. In order to answer this question, he divided the topic into two parts:

(a) a situation in which costs increase in response to short-period increases in demand but fall in response to long-period increases in demand;
(b) a situation where costs fall also in response to short-period increases in demand.

In case (a) the firm works to capacity (or even above capacity) but its size is smaller than the optimum scale; in case (b) the firm is not working to capacity when an increase in demand occurs.

Case (a) can take place in situations in which the rate of expansion of the optimum source of supply is larger than the rate of expansion of demand. In this case, the decrease in cost takes place in response to long-period increases in demand. These industries can be considered 'increasing returns industries' (Harrod 1931:99). As to case (b), Harrod justified the possibility of decreasing short-period costs in response to a change in demand by holding that

> If the prospective normal demand of the equilibrium firm is such that it will not absorb the optimum output of the optimum plant, it is probable that the firm will construct a plant, the optimum output of which will exceed the prospective normal output. In this case the plant, when producing for the normal demand, will show decreasing costs. If a firm is considering the desirability of reconstruction and the proper scale of operations, the question which it asks is, not—What is the plant the optimum output of which the normal demand will absorb? but—What is the plant with which the normal demandcan be met most cheaply? *If an*

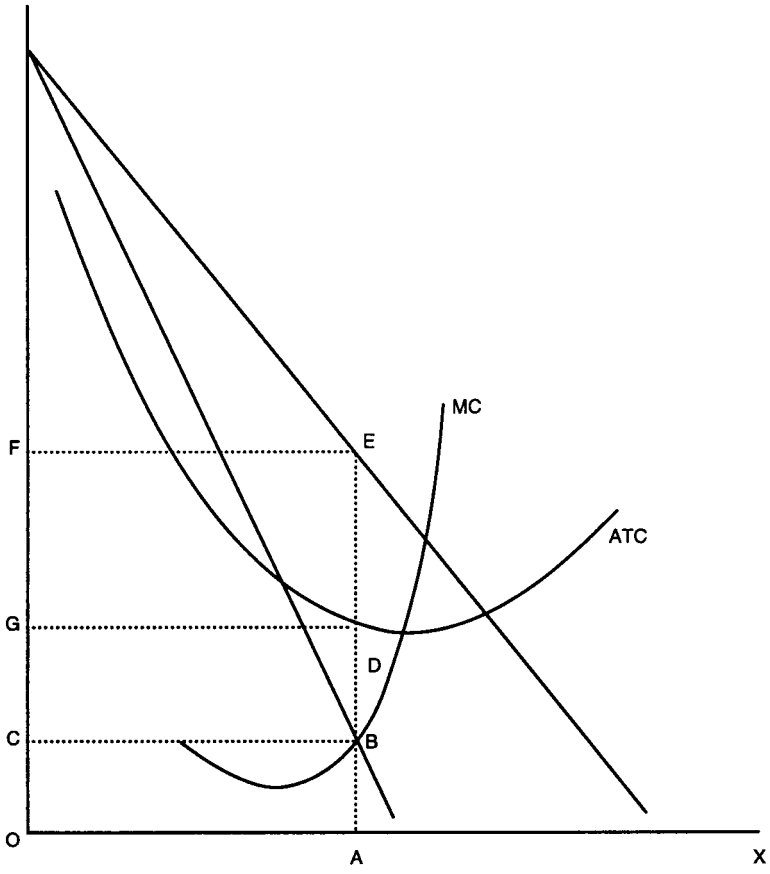

Figure 14.2 Marginal cost, marginal revenue and equilibrium

> *increase of scale provides substantial economies, such an increase may be desirable, even if full advantage of the economies cannot be taken.*
>
> (Harrod 1931:99; emphasis added)

Therefore, in normal times, a firm can experience decreasing costs also in response to a short-period increase in demand.

Harrod summarised his results in four points:

1 Competitive equilibrium is consistent with decreasing costs if there exist marketing expenses or the market is not perfectly unified, as in both cases the demand curve must be downward sloping;
2 profit is consistent with such equilibrium;

3 equilibrium is consistent with decreasing costs also in the long period if there are considerable transport costs or a strong spirit of individualism among entrepreneurs;
4 conditions of decreasing costs are normal in industries where the rate of expansion of the optimum plant is higher than the rate of expansion of demand.

Harrod's notion of equilibrium was a 'Marshallian evolutionary equilibrium and not a timeless solution of unchanging factors' (Shackle 1967:40). He introduced dynamic factors like the rate of technical change and the rate of expansion of demand in order to give an explanation of the existence of excess capacity in the long period. Robinson was soon to expunge these dynamic elements from the analysis of the problem of excess capacity.

ROBINSON'S SOLUTION OF THE EXCESS CAPACITY PROBLEM

In 1932, Robinson, in the article 'Imperfect competition and falling supply price',[6] dealt with the determination of equilibrium in imperfect competition, which, for her, had to be necessarily associated with firms experiencing excess capacity. She set her analysis within a framework that was rigorously static: time required for reaching equilibrium is ignored; firms are supposed to be always in individual equilibrium; the supply of every factor of production to the industry is perfectly elastic; there are no economies of large-scale industry; all firms are similar with respect to costs and conditions of demand, even though they are not alike from the viewpoint of buyers lest the market should be perfect (Robinson 1932:544 – 5). In order to avoid dealing with marketing expenses, imperfections are regarded as generated only by differences in transport costs.

In her article, Robinson used the term 'marginal revenue' for the first time. The individual firm maximises its profit when marginal revenue *MR* equals marginal cost *MC*. When *MC=MR,* the firm is not induced to vary its output even though there is no reason why the equilibrium output so determined is the optimum (full capacity) level the firm could produce. The equilibrium of individual firms does not imply that the industry as a whole is in equilibrium. In the long period, new firms could enter the industry and make aggregate output vary. The industry will reach equilibrium when profits are normal (Robinson 1932:546).

Robinson's notion of industry equilibrium raises two issues that are strictly related to one another: is it legitimate to consider free entry as compatible with imperfect competition? How is normal profit to be defined? At first sight, free entry appears to be incompatible with imperfect competition: imperfections themselves make the entry of new firms difficult. Inother words, under imperfect competition, the profit that a firm must expect in order to be induced to enter an

imperfect market has to be higher than under perfect competition. This, however, did not mean for Robinson that the profit had to be higher than normal. Normal profit is that profit which is just sufficient to dissuade the entry of new firms into a certain market (Robinson 1932:547). Robinson was aware of the abstractness and artificiality of this approach[7] but she followed it in order to make her analysis simpler.

A normal profit so defined is then included in firms' long-period costs. The industry is in equilibrium when the price equals the average total cost, i.e. the average revenue is equal to the average total cost:

> The equilibrium of the industry thus requires a double condition. Marginal revenue must be equal to marginal cost, and price must be equal to average cost. This double condition of equilibrium can be fulfilled *only when the individual demand curve of the firm is a tangent to its average cost*.
>
> (Robinson 1932:547, emphasis added)

Robinson ascribed this result to Kahn. The result was called 'Mr Kahn's theorem' by Shove (1933). We can now draw the graph of an industry equilibrium in which firms experience excess capacity (Figure 14.3).

At the intersection of *MC* and *MR,* the *MC* curve can be decreasing as in Figure 14.3, but it could also be either increasing or constant.[8] In any case, *at the equilibrium point E, the average total cost curve must necessarily be decreasing* since it has to be a tangent to the demand curve, which has a negative gradient.

As to Harrod's attempt to justify excess capacity with dynamic analysis, Robinson objected that there was no need to abandon a purely static framework to get to such result:

> He resorts, for an explanation of falling average costs, to the fact that the optimum size of a firm…may alter faster than the firm can grow…. This is a point of great interest, but it is not necessary to the matter. The firm will be of less than optimum size, in an imperfect market, however long the optimum remains unchanged.
>
> (Robinson 1932:549n)

THE ENSUING DISCUSSION: HARROD AND ROBINSON

Shove and Harrod soon questioned Robinson's 1932 result. Shove (1933) did not deal directly with the issue of excess capacity, but his criticism ofRobinson's notion of equilibrium has implications for the doctrine of excess capacity. Shove' argument was taken up by Harrod to reject the conclusion that, in imperfect competition, equilibrium is *necessarily* associated with excess capacity.

As we saw, for Robinson, an industry is in equilibrium when firms earn a normal profit and there is no tendency for the aggregate output to change. Shove

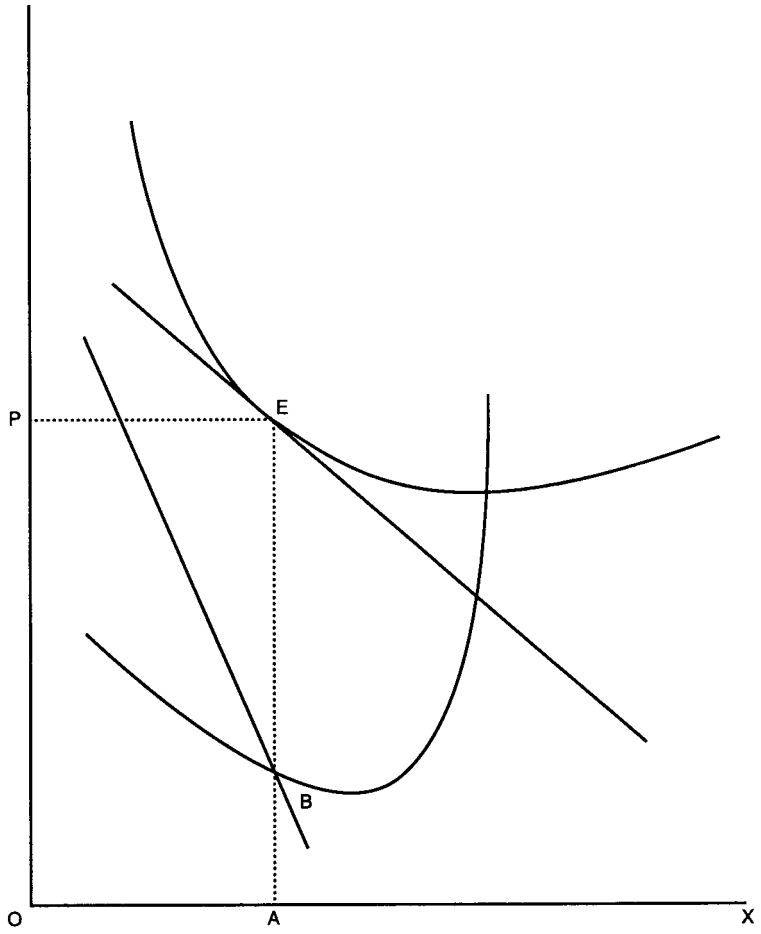

Figure 14.3 An industry equilibrium in which firms experience excess capacity

argued that the industry equilibrium, interpreted in the sense that there is no tendency for the output to change, does not require profits to be normal: they can be below the normal level provided that they are above that level which would cause some already existing firms to leave theindustry (Shove 1933:119). If there were perfect competition and hence no costs of entry into an industry, the two limits for profits would coincide, but in imperfect competition these costs are positive, so that the two levels do not coincide.

Thus, in imperfect competition, although profits are below their normal level, the industry output does not vary. This means that the stability of the industry output does not necessarily imply that the price must be equal to the average

cost. The equality between price and average cost is necessary only to ensure the maximum level of profits. Therefore

> Mr Kahn's theorem does not mean that the individual demand curve of the firm must be tangent to its average cost curve in order that there should be no tendency for the industry's output to alter, but merely in order that the firm should be earning this maximum profit.
>
> (Shove 1933:121)

Robinson maintained and even reinforced her line of approach in *The Economics of Imperfect Competition* (1933a). Here she did not substantially change her previous position on equilibrium and normal profits, but some further qualifications were introduced. Three different periods of time were used: the *short period* (the firm's equipment is fixed); the *quasi-long period* (the firm's equipment can vary to adapt to changes in output); the *long period* (also the number of firms can vary) (Robinson 1933a:47). As to the long period, the analysis was essentially the same as that developed in 1932, but the unrealism of most hypotheses was more stressed, especially in footnotes. The level of profits in the industry governs the entry or exit of firms; when the profit is at its normal level the number of firms in an industry remains constant. Normal profit is defined in the same way as in 1932. Firms realise normal profits at the point of tangency between average total cost and demand curve, so that in full equilibrium[9] firms have a certain degree of unused capacity (Robinson 1933a: 97).

Robinson's distinction between quasi-long period and long period is useful to understand better Harrod's reasoning in his 1933 and 1934 contributions to the debate. In 1933, Harrod referred only to equilibrium in the quasi-long period, even though he adopted a different terminology. Harrod took up Shove's point concerning normal profits, and held that Robinson's definition of normal rate of profit did not specify a unique rate, but a range of rates.[10] However, a unique normal rate of profit is necessary to draw average total cost curves. Thus Harrod looked for an alternative definition of normal rate of profit and proposed that rate which is 'just sufficient to induce the *firm itself* to embark on additional fixed capital expenditure' (Harrod 1933a:104). This rate is presumably lower than the rate that wouldattract new firms into the industry and higher than the rate that would induce the firm to leave the industry.

With this new definition of normal rate, 'Mr Kahn's theorem' is no longer necessarily true. In short-period equilibrium, marginal cost and marginal revenue must be equal while price and average cost can be different; if they are equal the price curve and the cost curve must be tangential. In the long period, profits must lie within the lower and upper limits and, moreover, the firm will choose that amount of fixed equipment that allows it to produce at the lowest cost, which includes the normal rate of profit as defined by Harrod.[11] Therefore, in quasi-

long period equilibrium, 'the firm will be earning on its marginal outlay on fixed plant a rate of return which it regards as justifying fixed capital expenditure' (Harrod 1933a:105).

In quasi-long period equilibrium, the long-period cost curve and the curve showing total cost per unit of producing with the particular plant in use are tangential; this, however, does not necessarily imply that the two tangential cost curves are also tangential to the demand curve. If this is the case, it means that the equilibrium price ensures an average rate of profit on fixed capital which is equal to the rate used to construct the cost curve associated with that amount of fixed capital; in other words the average rate of profit is the same as the normal rate of profit and the marginal rate of profit. In this case, the firm also experiences excess capacity, i.e. a decreasing average cost.[12]

Harrod objected to the fact that a firm in equilibrium must necessarily earn a normal rate of profit on *all* its fixed equipment. It must earn a normal rate (as defined by him) only on its marginal fixed equipment (Harrod 1933a:106). This implies that it is not necessarily true that the demand curve is tangential to the long-period average cost curve. What is necessary is only that, at the equilibrium point, the (long-period) marginal cost curve intersects the marginal revenue curve, so that the marginal revenue to the firm from the marginal fixed equipment ensures a normal rate of return. In this way, quasi-long period equilibrium can be depicted with a diagram which is analogous to that used for the short-period equilibrium (see Figure 14.2). In other words,

> it is *not* a necessary condition of [quasi] long-period equilibrium that the price should be equal to the total average cost per unit. If the price is greater than the total average cost per unit, the gradient of the demand curve has a greater negative value than that of the total cost per unit curve; in this condition, therefore, the long-period cost curve may not have a negative gradient at all. If the price is equal to or less than the total average cost per unit, the long-period cost curve must have a negative gradient.
>
> (Harrod 1933a:106)[13]

Thus, for Harrod, in equilibrium under a regime of imperfect competition, a firm might, but it *must not necessarily,* experience long-period and short-period decreasing costs (Harrod 1933a:106).

In 1934, after the publication of Robinson's *The Economics of Imperfect Competition* and Chamberlin's *The Theory of Monopolistic Competition,* Harrod published another article, where his previous results in the theory of imperfect competition were restated and further developed also in the light of Robinson's and Chamberlin's contributions. Here, Harrod analysed also long-period equilibrium. Like in his previous works, he pointed out the necessity to establish a normal rate of profit in order to draw meaningful cost curves; as in 1933 he chose to adopt as the normal rate that rate which is the lowest that represents an

adequate inducement for the firm to invest in fixed equipment.[14] Harrod turned to 'true' long-period analysis by asking whether, in imperfect competition, excess profits earned by a firm already in an industry are eliminated by the entry of new firms. He considered a representative firm in imperfect competition, that by 'mere accident' is of the optimum size, produces to capacity and earns excess profits. This firm, of course, is not induced to alter its production: it does not earn any excess profit on its marginal fixed equipment (Harrod 1934:125). But will not the excess profits attract new firms? For Harrod, the answer to this question depended on the details of the market structure that makes competition imperfect (Harrod 1934:125).

In imperfect competition, a new competitor could not acquire a large part of the existing market by selling at the same price as the already existing firm. Therefore it should sell at a substantially lower price, and the reduction in the price is the larger the larger is the market share that the competitor wants to take from the existing firm. Moreover, if the competitor wants to reach its optimum scale, or even just to draw benefits from the economies of large scale, it has to acquire a large share of the market. In this situation, 'new-comers *may* be excluded, even if existing producers are making excess profits' (Harrod 1934: 125). Thus excess profits are a necessary corollary of the adoption of optimum methods of production in imperfect competition (Harrod 1934:125). Equilibrium with excess capacity,

> would represent long-period equilibrium in an industry on the *unlikely condition* that, in spite of imperfect competition, new competitors can enter the business and find a market as easily as pre-existent sources. In that event profits would tend to a standard level.
>
> (Harrod 1934:126, emphasis added)

In other words, equilibrium of the sort hypothesised by Robinson, though possible, is unlikely.

Harrod's criticism of Robinson in 1934 rested on regarding as unlikelythe hypothesis that in imperfect competition, in the same way as in perfect competition, free entry eliminates excess profits.[15] If the existence of obstacles to entry, a typical phenomenon in imperfect competition, prevents the elimination of excess profits, it is no longer necessarily true that firms must be of less than optimum size and work below capacity. Equilibrium in imperfect competition is not necessarily associated with excess capacity. Harrod summarised his position in the following way:

> In conditions of imperfect competition, long-period equilibrium is consistent with lay-outs of less than the optimum size and their exploitation below their optimum capacity. If not more than standard profits are earned in equilibrium, the lay-out *must* be too small and it *must* be worked below

> capacity. If, on the other hand, optimum conditions of production are achieved, profits *must* be above the standard level.
>
> (Harrod 1934:127 – 8)

We can express Harrod's point of view also in the following way: if there is not free entry into an industry, the number of operating firms is likely to be smaller than in the case of free entry. For the smaller number of operating firms it is then possible to produce in optimal conditions and still earn profits above normal. On the other hand, if there were free entry, the number of producers would tend to be too large with respect to demand, and all producers would be prevented from expanding their size and output to the optimum level.

Robinson commented on Harrod's criticisms in two articles (Robinson 1933b, 1934). She argued that it was not correct to characterise imperfect competition by the lack of free entry. In her short 1933 note, Robinson recognised the necessity to distinguish between the level of profit that prevents new firms from entering an industry and the level of profit that does not induce existing firms to leave an industry. However, she held that such a gap between the two levels of profit was not connected with the form of the market (Robinson 1933b:531).[16] With this qualification, Robinson did not see any difficulty with her notion of normal profit, while Harrod's definition was considered 'extremely obscure'.

Robinson's 1934 article (Robinson 1934) was mainly devoted to the definition of perfect competition, and in this context she argued again that the notion of perfect competition does not have to be associated with free entry. The two notions that, in perfect competition, the demand curves faced by single firms are perfectly elastic and there is free entry, are usually 'lumped together', so that no individual seller can earn more than normal profits. But, for Robinson, these two aspects must be disentangled and the definition of perfect competition must be restricted to the existence of perfectly elastic individual demand curves (Robinson 1934:20 – 1). Thus, that there must be anormal profit is true also in an imperfect market. The distinction between perfect and imperfect competition reduces to the different elasticities of the demand curves faced by single sellers.

By arguing that barriers to entry need not characterise imperfect competition, Robinson was again holding, though indirectly, that 'Mr Kahn's theorem' had a general validity. However, after Harrod's criticisms, she had to state explicitly that imperfect competition is not distinguished from perfect competition from the point of view of free entry. This was a position that would be further questioned in the years to follow.

THE ENSUING DISCUSSION: CHAMBERLIN, HARROD AND KALDOR

Chamberlin's (1933) book came out in the same year as Robinson's, although the origins of his work were different.[17] However, as far as equilibrium with

excess capacity was concerned, Chamberlin's results were essentially the same: a firm in imperfect competition maximises its profit when marginal cost is equal to marginal revenue. In equilibrium, the firm is not induced either to change its price, the quantity supplied or the product. In equilibrium there is no reason why the firm should produce to capacity, and its profit can be above the minimum. In order to see how profits are, or can be, reduced to their minimum it is necessary to take into consideration the possibility for other firms to enter the 'group' and carry out the analysis of group equilibrium.

The group—the set of producers which constitute an imperfectly competitive market (e.g. automobile producers, shoe producers, etc.)—by its own nature, is characterised by a large variety of prices, differentiated outputs, scales of production and profits, so that 'the imperfection of competition is not uniform'. However, after a number of reservations and specifications, Chamberlin considered it possible to make the 'heroic' assumption that both demand and cost curves for all products were uniform. Moreover, he assumed that the number of producers belonging to the group was large, so that individual changes in price and/or product give rise to effects that are so widely spread among competitors to be negligible (Chamberlin 1933:83).

The existence of extra profits attracts new firms into the group. Their entry makes the demand curve for each producer's product shift to the left, and the process of entry will continue until the demand curves have reached such a position that no extra profits can be earned by any producer, that is to say until the cost curve is tangential to the demand curve for each producer. Since demand curves are downward sloping, the tangency position can be realised only at a point where the firm produces at decreasing costs, i.e. below capacity. Chamberlin examined producers' behaviour in moredetail than Robinson did, but he arrived at the same result as to equilibrium with excess capacity.

The conclusion above concerning excess capacity was reached under the assumption that producers do not take into account the indirect effect of their actions. Chamberlin, however, considered also the case in which rivals' reactions are taken into account. He first analysed equilibrium in a small group without product differentiation (Chamberlin 1933:30 – 55) and concluded that the equilibrium price realised in the group depends on the kind of producer behaviour that is assumed. In particular, if producers take account of both direct and indirect effects of their behaviour, the equilibrium price will be the same as the equilibrium price in monopoly.[18] Both when producers first adjust quantities and when they first adjust prices, there is no tendency for the equilibrium price to reach the same level as in a perfectly competitive market. The tendency to the equilibrium competitive price can become dominant only when the effects are negligible (Chamberlin 1993:54). In examining equilibrium of a small group with product differentiation, Chamberlin arrived at a similar conclusion: equilibrium of a small group does not necessarily imply that extra profits are eliminated. In particular, if producers take account of the total (i.e. direct and

indirect) effect of their actions, extra profits are at their maximum (Chamberlin 1933: 101 – 2).[19] However, for Chamberlin, these results did not also imply that profits must be above their minimum. Entry of new firms into the group will bring profits down to their minimum, that is, it will bring demand curves to be tangent to average cost curves, even if the price does not change and remains at the same level as it was before the entry of new competitors.[20]

Chamberlin then dealt with the assumption that cost and demand curves are identical for all firms in the group (Chamberlin 1933:110 – 13). He first considered cases where curves differ from one another in their position or in shape. In both cases free entry would bring each producer to the 'tangency position', so that each firm produces below capacity. More interesting is the third case, where curves differ for their relative position to one another. Up to this point, Chamberlin had argued that the entry of new competitors would have brought the cost and demand curves to be tangent; this conclusion was based on the assumption that it was possible to produce substitutes which were able to compete with the already existing products. However,

> in so far as substitutes of such a degree of effectiveness may not be produced, the conclusions are different—demand curves will lie to the right of the point of tangency with cost curves, and profits will be correspondingly higher. This is the explanation of *all* monopoly profits, of whatever sort.
>
> (Chamberlin 1933:111)

In other words, if there is not perfectly free entry excess profits may not disappear. Chamberlin concentrated on monopoly profits, but he failed to follow through his argument and point out that, in such a situation, it could be true that some (or all) firms in the group produce to capacity: if the cost and demand curves must no longer be necessarily tangent in equilibrium, firms could, at least in principle, produce a level of output corresponding to their capacity output.

Finally, Chamberlin also analysed the problem of excess capacity by taking account of marketing. When expenses for advertising ('selling costs') are considered, the conclusions concerning excess capacity can be partly modified. Selling costs are borne by firms in order to increase the demand for their products; i.e. to shift the curve to the right, but at the same time advertising also makes cost curves shift to the right because they now include selling costs. In this situation it is possible, although unlikely, that the larger demand for firms' products determines a higher level of output than would have been produced in the absence of selling costs (Chamberlin 1933:171 – 2). Excess capacity is necessarily reduced if selling costs do not modify, or increase, the elasticity of the demand curve. If, instead, elasticity of demand is reduced, firms produce less than in absence of selling costs.

In a review of *The Theory of Monopolistic Competition,* Harrod (1933b), after having criticised Chamberlin's solution of equilibrium of a small group with no

product differentiation,[21] turned to consider the problem of the tendency of profits toward their minimum level in monopolistic competition (Harrod 1933b: 662 – 3). He asked the question why existing firms, which behave oligopolistically, should not carry out price policies aimed at preventing new firms from entering the market. For Harrod, the hypothesis of free entry had to be questioned or, at least, the hypothesis of barriers to entry should receive the same amount of attention and analysis as the hypothesis of free entry.[22] Harrod criticised Chamberlin from the same perspective as that which he had adopted to criticise Robinson.

In 1935, Kaldor, in examining the problem of excess capacity in imperfect competition, criticised Chamberlin from a similar point of view.[23] In his critique of the 'doctrine of excess capacity', Kaldor referred only to Chamberlin's work because he regarded Robinson's results as vitiated by tautology.[24] Kaldor summarised the basic assumptions on which Chamberlin's results on excess capacity were based, and formulated the final result in the following way. If in the industry there are initially extra profits,

> new firms will come in…and this process will continue, until profits are reduced to normal, i.e. the difference between the actual earnings and the displacement costs of the entrepreneur's own resources is eliminated.[25] In the position of final equilibrium…[t]he demand curve will thus be tangential to the cost curve…. The producers, *as a body,* could of course prevent this from occurring by reducing theirprice *in anticipation* of the entry of new competitors. But since the appearance of any *single* new producer will only affect the demand for a *single* existing producer very slightly, while similarly the reduction of price of a *single* existing producer will only slightly affect the profits which a potential producer can expect, no producer could take these indirect effects on his own price policy into consideration. There can be little doubt that given these assumptions the theory is unassailable.
>
> (Kaldor 1935:65 – 6)

Thus criticisms must be addressed to the assumptions, on which that theory was based. Kaldor first stressed the fact that also large groups can be equated to small groups (Kaldor 1935:67 – 70), a point that, as we have seen, had already been made by Chamberlin himself. But Kaldor went further and developed what Harrod too had pointed out in his review of Chamberlin's book: it is legitimate to think that producers also take into account the reactions of potential competitors.

> If a producer knows that if he charges a high price to-day a competitor will appear to-morrow whose mere existence will put him in a *permanently worse position,* he will charge a price which will afford him only a low profit, if only he hopes to secure this profit permanently; i.e. he will act in a manner

> *as if* his own demand curve were very much more elastic than it is. And this 'foresight' will, or at any rate may, prevent him from being driven to a state of excess capacity.
>
> (Kaldor 1935:70)

Kaldor also showed that, if there are economies of scale, competition of new entrants would not be such as to make demand and cost curves tangential, even if producers do not take into account indirect effects of their actions. In the presence of economies of scale, and hence of indivisibilities, the entry of new firms will not necessarily continue until the 'tangency position' is reached. The basic reason for this is that the potential competitor, just because of economies of scale, may be prevented from entering the market because they know that they could enter the market efficiently only by producing such an amount of output that would cause a fall in the price, which gives rise to a loss rather than to a profit.[26]

Kaldor then turned to consider another aspect that Chamberlin had already considered, albeit marginally: the possibility that cost and demand curves differ from one firm to another. He held that maintaining the assumption that all cost and demand curves are similar in all firms could be justified only if it were also assumed that there were no institutional monopolies.[27] But institutional monopolies are never completely absent, andtherefore it is possible that extra profits cannot be competed away (Kaldor 1935:74), a point already made by Chamberlin even though without emphasising the implications in terms of excess capacity. Finally, Kaldor took into consideration the implications of not considering joint production when dealing with the issue of excess capacity. If there is the possibility of joint production, excess capacity could be eliminated by firms by producing other goods.

CONCLUSION

The origin of the doctrine of excess capacity was an attempt to fill the gap between the theoretical outcomes of the then-prevailing theory of the firm and the actual behaviour of firms during the depression of the 1930s. In the preface to the second edition of *The Economics of Imperfect Competition* (1969), Robinson provided a short, though illuminating, reconstruction of the process that led to the development of 'Mr Kahn's theorem'. Pigou transformed Marshall's rather vague notion of competition into a 'neat and logical system'. Within Pigou's system, firms always tend either to produce the level of output associated with their optimum size, or not to produce at all (when the price is lower than the average cost). Such a conclusion was evidently at odds with the reality of the 1930s. The notion of imperfect competition seemed to allow economists to offer a better explanation of the actual behaviour of firms and markets during a depression.

> Under perfect competition, any plant that was working at all must be working to capacity…. Imperfect competition came in to explain the fact, in the world around us, that more or less all plants were working part time…. The notion that every firm is facing a falling demand curve… provided an explanation for a situation in which firms could work their plants at less than full capacity and still earn a profit.
>
> (Robinson 1969:vi)

Unfortunately, Robinson's results did not prove to be much more realistic than the traditional ones. Her demonstration of non-competitive equilibria with excess capacity was based on hypotheses on the working of markets that were largely unsatisfactory in terms of realism. Furthermore, she rejected Harrod's earlier attempt to enrich the analysis with dynamic considerations, and chose a purely static approach to the problem of excess capacity. Chamberlin was more cautious, and his analysis was more complex and articulated than was Robinson's, but he too had to make extremely unrealistic hypotheses in order to achieve the result of equilibrium with excess capacity.

Harrod's and Kaldor's criticisms of the doctrine of excess capacity were all based on the need for more realistic hypotheses in the analysis of firms working in a non-perfect competitive environment. Both questioned the hypothesis of free entry that was so crucial for Robinson's and Chamberlin's results. Why should monopolistic competitors behave in such a way as to leave potential competitors free to invade their own markets? It was both more realistic and logical to think that firms' policies in imperfect competition are such that there arise relevant barriers to entry. If such barriers exist, there is no necessary tendency for extra profits to disappear, and then it is no longer necessarily true that firms must experience excess capacity. Robinson's reply to these criticisms was not convincing. She tried to argue that facility of entry into an industry had nothing to do with the degree of competition. A very peculiar position, since it seems that the existence of any non-perfect market form must be necessarily contingent on the existence of some form or another of obstacles to the free movement of factors.

It is interesting that both Harrod and Kaldor referred to Shove's criticism of Robinson's treatment of the topics in discussion. As we saw, Shove did not enter into the specific discussion of excess capacity, but he provided some basic elements for the rejection of 'Mr Kahn's theorem'. At a more general level, Shove argued that, if some realism of hypotheses had to be retained, the analysis of the working of markets should have been carried out by renouncing formal rigour.

> So long as we are content with a rough and ready indication of the forces at work, we can keep fairly near to the facts: but any attempt to make our treatment exact is apt to lead either to a degree of abstraction which

> renders the apparatus inapplicable to the actual phenomena we set out to explain or to a degree of complication which makes it cumbrous to use.
>
> (Shove 1933:121)

It is this scepticism about the possibility of developing a formally rigorous analysis of the microeconomy that may be a clue to understanding also Keynes' attitude toward the debate on imperfect competition and excess capacity. As is well known, in *The General Theory,* Keynes completely ignored this debate. This, however, was certainly not due to his ignorance of the topic. Although not deeply involved, he was aware of the discussion on market forms and the problem of excess capacity.

In general, Keynes did not see any significant relationship between the issue of market forms and his principle of effective demand.[28] As to the doctrine of excess capacity in particular, his conviction was that it was far from satisfactory. Indicative of his attitude is a piece of correspondence with Robinson on an article that Kalecki, in 1941, had submitted for publication in *The Economic Journal*. Kalecki assumed that firms always work belowcapacity, and Keynes observed: 'Is it not rather odd when dealing with "long run problems" to start with the assumption that all firms are always working below capacity?' (Keynes 1983:829). For Joan Robinson, it was not at all odd, because that was 'part of the usual bag of tricks of Imperfect Competition Theory' (cited in Keynes 1983: 830). Keynes replied:

> If he [Kalecki] is extending the *General Theory* beyond the short period but not to the long period in the old sense, he really must tell us what the sense is. For I am still innocent enough to be bewildered by the idea that the assumption of all firms always working below capacity is consistent with a 'long-run problem'. To tell me that as for under-capacity working that is part of the usual pack of tricks of imperfect competition theory does not carry me any further. For publication in the *Journal* an article must pass beyond the stage of esoteric abracadabra.
>
> (Keynes 1983:830 – 1)

He added that Kalecki was 'taking artificial assumptions which have no possible relation to reality or any other merit except that they happen to lead up to a needed result'.

Keynes's scepticism about the 'bag of tricks' of imperfect competition, on the other hand, was accompanied by the conviction that microeconomics should develop along a similar line to that which Shove had been trying to follow. An exchange of correspondence with Shove is revealing. A few weeks after the publication of *The General Theory,* Shove wrote to Keynes:

> I thought you were too kind to the 'classical' analysis as applied to the individual industry and firm. Unless very special assumptions…are made, it seems to me either wrong or completely jejeune. I have been groping all these years after a re-statement of it on lines similar in some respects to your solution for the system as a whole, stressing in particular 'expectations' and the influence of current and immediately past experience upon them. But I can't make it precise.
>
> (cited in Keynes 1973:1)

Keynes replied:

> What you say about the Classical analysis as applied to the individual industry and firm is probably right. I have been concentrating on the other problem, and have not, like you, thought very much about the elements of the system. But you ought not to feel inhibited by a difficulty in making the solution precise. It may be that a part of the error in the Classical analysis is due to thatattempt. As soon as one is dealing with the influence of expectations and of transitory experience, one is, in the nature of things, outside the realm of the formally exact.
>
> (Keynes 1973:2)

The 'neat and precise' results of the doctrine of excess capacity would not fit into Keynes' approach to the analysis of economic phenomena.

Moreover, even if we ignore the methodological considerations above and remain at a mere analytical level, the existence of excess capacity equilibria does not have, *per se,* any direct implications for the analysis of aggregate unemployment. All industries might experience a certain degree of underutilisation of capacity and there could still be full employment of labour. On the other hand, also in a situation of unused capacity and labour, the Keynesian recipe for increasing the level of employment would be the same as under a competitive hypothesis. Whatever the prevailing market form is, an increase in employment requires an increase in effective demand.

There is only one way in which a hypothesis of firms experiencing excess capacity could have significant implications for Keynes' theory. The existence of unused capacity would be likely to affect firms' investment decisions. It would represent a constraint to their decisions concerning expansion of their size. If firms' investment is constrained by a significant degree of under-utilisation of existing plant, this has in turn implications for the level of aggregate demand, and hence for the level of aggregate employment. It is unfortunate that Keynes did not develop this analytical aspect. Certainly, he was not helped to do so by the advocates of the doctrine of excess capacity, who restrained their analysis within a purely static and unrealistic context without paying any attention to the expansion of firms over time. The critics of that doctrine, on the other hand, were

not able to offer a sufficiently solid alternative approach that could have favoured the development of a higher integration between micro and macro theory.

NOTES

1 Although he certainly exerted a great influence on Joan Robinson, one of the most prominent participants in the debate.
2 The other three methods are: altering the intensity of use of machinery, altering the number of working hours per shift, and altering the number of shifts (Kahn 1989: 46).
3 For Kahn, prime costs usually include some elements which have an overhead nature because they are inflexible (e.g. costs for fuel, lighting, repairs, etc.). They rise steeply when production passes from zero to a low level, and then increase very little for higher levels of output.
4 See Shackle (1967:22 – 42) for a history of the notion of marginal revenue.
5 Harrod also observed that increasing marketing costs could be due to the requirement of a more intensive salesmanship as output expands. A more intensive salesmanship could be regarded as insufficient to ensure a long-period equilibrium only if it could be assumed that firms are willing to lose their individual identity and tend to concentration. But for Harrod, firms' 'individualism' has to be considered an influence of lasting importance, so that the 'period of the equilibrium maintained by the presence of these kinds of costs is a sufficiently long one' (Harrod 1931:98).
6 The main object of the article was to show that decreasing average costs do not necessarily imply decreasing supply prices.
7 'An increase in demand attracts new entrepreneurs to the industry directly, by opening up some new possibility of profitable investment, rather than indirectly, by making their mouth water at the sight of the high profits of the existing firms. The abnormal profits are a symptom rather than a cause of the situation in which new firms will find it profitable to enter the trade' (Robinson 1932:547n).
8 In order that there be an equilibrium, it has to be, of course, $dM\ R/dx<dM\ C/dx$ where x is the output.
9 An industry was said to be in *full equilibrium* when there is no tendency to changes in the number of firms.
10 '[I]f there were a unique rate such that any rate in excess of it attracted new competitors and any rate below it caused defection, we should surely be in the sphere of perfect competition' (Harrod 1972:103).
11 Each cost curve of those that lie tangential to the long-period cost curve (i.e. their envelope) includes such a normal rate of profit.
12 'If, as Mrs Robinson implies, the firm must earn a unique normal rate of profit on all its fixed equipment in long-period equilibrium, then her conclusion follows that the total cost per unit curve, which embodies this rate, must be tangential to the demand curve, and, since that has a negative gradient in an imperfect market, an equilibrium firm working for an imperfect market must show decreasing costs'(Harrod 1972:105 – 6).

13 If the price is higher than the average cost, the firm's average rate of profit is higher than its marginal (normal) rate of profit; if the price is lower than the average cost, the average rate of profit is lower than the marginal rate of profit.
14 In 1934, this rate was called *standard* rate of profit, instead of normal (Harrod 1934:118).
15 It remains true, however, that the *marginal* rate of profit must be normal (or standard) because, otherwise, either existing firms would be induced to expand their output or new firms would enter the industry.
16 When an expanding industry (in perfect or imperfect competition) is considered, the normal profit is represented by its upper level; when a contracting industry is considered, the normal profit is represented by its lower limit.
17 Chamberlin has always held that the concept of monopolistic competition is different from that of imperfect competition.
18 When producers do not take account of indirect effects of their actions, i.e. when they assume their rivals do not react, we can have either a solution a la Cournot or one a la Bertrand (Chamberlin 1933:32 – 46).
19 Chamberlin's analysis of a small group is relevant because he pointed out that a large group could be considered as made up of several small groups. In fact, any single firm is actually in close competition only with a small number of other firms, so that it is not legitimate to think that the effects of the firm's actions on its competitors are small and hence negligible. In such a case, the firm considers the total effects of its actions, and it might avoid price competition because it expects that others will follow suit (Chamberlin 1933:103). Thus the general conclusion is that 'considerations relevant to competition between small numbersare much more generally applicable than might at first be supposed' (Chamberlin 1933:104).
20 A situation like the following could occur. Before the entry the price p ensures an extra profit for each producer supplying the quantity $X;$ after the entry of new competitors, the demand curve shifts to the left, each producer's supplies $X'<X$ and does not earn any excess profit even though the equilibrium price is still at the same level as before.
21 Harrod criticised Chamberlin's idea that there is a 'break' between a situation where producers take account of the effects of their actions and a situation where they ignore such effects because they have become negligible. For Harrod, there was no such a break but a gradual transition from a monopolist to a competitive solution (Harrod 1933b:662 – 3).
22 Harrod in fact recognised that Chamberlin was aware of the differences that lack of free entry would imply; but he complained that Chamberlin did not develop this aspect (see Harrod 1933b:665).
23 Although Kaldor was less generous than Harrod in recognising that many of his points had already been mentioned by Chamberlin without stressing their importance.
24 'Mrs. Robinson's version possesses a merely formal similarity with Professor Chamberlin's theory. For Mrs. Robinson includes in her "cost curves" such profits which are not competed away by the entry of new producers; and in the circumstances, her statement that "demand curves will be tangential to cost curves" and that firms will be of "less than their optimum size" is merely a statement of a tautology' (Kaldor 1935:63). On this point, Kaldor also referred to Shove's 1933

critique of Robinson. Kaldor thought that Shove's article contained 'one of the most penetrating analyses so far published on this whole subject' (Kaldor 1935:63n).

25 Kaldor included the displacement cost of entrepreneurs' resources in the average cost curve.

26 Kaldor acknowledged that his conclusions would not hold under the assumption that consumers' preferences are evenly distributed, so that the entry of a new firm affects all the existing ones and the reduction in demand is negligible. But he also stressed the unrealistic nature of such an assumption (Kaldor 1935:72).

27 In this respect, Kaldor observed: 'Professor Chamberlin does not state this explicitly; but this is the only logically consistent interpretation one can give to his assumption that "the entry of new producers into the field in general and every portion of it in particular is free and unimpeded"' (Kaldor 1935:73n).

28 See for example Keynes' observations on Ohlin's comments regarding Robinson's doctrine of imperfect competition in relation to *The General Theory* (Keynes 1983: 190).

REFERENCES

Chamberlin, E.H. (1933) *The Theory of Monopolistic Competition: A Re-orientation of the Theory of Value,* 8th edn (1962) Harvard MA: Harvard University Press.

Harrod, R. (1930) 'Notes on supply', *Economic Journal,* June. Reprinted in Harrod (1972) *Economic Essays,* 2nd edn, London: Macmillan, 77 – 88.

——(1931) 'The law of decreasing costs', *Economic Journal,* December, 566 – 76.

——(1932) 'Decreasing costs: an addendum', *Economic Journal,* September, 490 – 2.

——(1933a) 'A further note on decreasing costs', *Economic Journal,* June. Reprinted in Harrod (1972) *Economic Essays,* 2nd edn, London: Macmillan, 103 – 7.

——(1933b) '*The Theory of Monopolistic Competition* by Edward Chamberlin', *Economic Journal,* December, 661 – 6.

——(1934) 'Doctrines of imperfect competition', *Quarterly Journal of Economics,* May. Reprinted in Harrod (1972) *Economic Essays,* 2nd edn, London: Macmillan, 111 – 38.

——(1972) *Economic Essays,* 2nd edn, London: Macmillan.

Kahn, R.F. (1929) [1989] *The Economics of the Short Period,* London: Macmillan.

Kaldor, N. (1935) 'Market imperfection and excess capacity', *Economica,* 2, 33 – 50. Reprinted in Kaldor (1960) *Essays on Value and Distribution,* London: Duckworth, 62 – 95.

——(1960) *Essays on Value and Distribution,* London: Duckworth.

Keynes, J.M. (1973) *Collected Writings,* vol. XIII, London: Macmillan.

——(1983) *Collected Writings* vol. XII, London: Macmillan.

Robinson, J.V. (1932) 'Imperfect competition and falling supply price', *Economic Journal,* December, 544 – 54.

——(1933a) *The Economics of Imperfect Competition,* 2nd edn (1969) London: Macmillan.

——(1933b) 'Decreasing costs: a reply to Mr Harrod', *Economic Journal,* September, 531 – 2.

——(1934) 'What is perfect competition?', *Quarterly Journal of Economics,* November. Reprinted in Robinson (1951) *Collected Economic Papers,* vol. I, Oxford: Blackwell, 20 – 34.

——(1951) *Collected Economic Papers,* vol. I, Oxford: Blackwell.

Shackle, G.L.S. (1967) *The Years of High Theory,* Cambridge: Cambridge University Press.

Shove, G.F (1930) 'The representative firm and increasing returns', *Economic Journal,* March, 94 – 116.

——(1933) 'The imperfection of the market: a further note', *Economic Journal,* March, 113 – 24, 92 – 3.

15
CANTABRIGIAN PREFERENCES
Joan Robinson v. Frank Hahn

Keiran Sharpe

> When we admit the influence of society, of the Joneses, of advertisement, upon the individual's scale of preferences, the problem of framing the [preference revealing] experiment becomes teasing indeed. Worse still, when we recognize that one man's consumption may reduce the welfare of others…we begin to doubt whether preferences are what we really prefer.
>
> (Joan Robinson 1962:51)

> Unlike Joan Robinson (1962), Kornai understands that the empirical content of preference theory is that preferences are relatively stable. He [and she] claims this to be false for all sorts of reasons such as the invention of new commodities, advertising and so on…. But when a good theorist thinks about preferences he is careful to describe the space over which they are taken to be defined. It is by now rather well known that we can think as easily of a 'quality space' as we can of a 'commodity space'.
>
> (Frank Hahn 1984:39)

In her little book, *Economic Philosophy,* Joan Robinson undertakes a commentary on utility (Chapter 3) which is partly a historical discursus and partly a critique of the concept. The discussion, as one would expect in a short personal tract covering the whole of political economy, is spare; and, since there is so little space to develop arguments, nothing much is made of the insights offered. Nevertheless, they contain all that is necessary to develop a thoroughgoing critical assessment of the axiomatic account of preference and choice; which is to say that her arguments can be read as a kind of précis of a thoroughgoing critique of neoclassical choice theory. This critique, it has to be emphasised, is a logical one, that is, it goes to the internal coherence of the axiomatic account; it has nothing to do with theempirical question as to whether or not preferences are stable over time, *pace* Hahn. In this regard, the critique bears a certain similarity with that made in the Capital Debates: a fact of which

Joan Robinson was herself aware (1962: 66 – 7).[1] That being said, it needs to be emphasised that her own presentation of the issues in Chapter 3 of her book is, as suggested, really only a 'prelude to a critique'. The purpose of this essay, then, is to sketch in some of the details of the fuller version. The general plan of attack is as follows: first, some counter-examples to the axiomatic theory of choice are presented which undercut its universalistic claims; the logical implications for the theory are then drawn; these negative conclusions are then canvassed in some explicitly economic contexts; and, finally, some implications for welfare economics are drawn.

Now, it is a basic principal of neoclassical economics that all consumers have well defined preferences over commodities. Commodities—the things that agents are after—are able to be measured and so can be taken as points in (commodity) space. The axiom set for preferences is such as to make the definition of a choice function over the domain possible. This choice function is equivalent to maximising utility, where the latter is an ordinal measure of preference. In the case of choice under uncertainty, other axioms (lottery reduction, Archimedean and independence-substitution) are super-added to arrive at a choice function which is equivalent to maximising expected utility. The utility in this case is (somewhat) cardinal (i.e. is unique up to a linear transformation) and the probability assessments are subjective (i.e. are measures of the probability at which an agent is indifferent between taking the bet constituted of the problem at hand versus some appropriately constituted 'ethically neutral' bet). In both the case of certain choice and choice under uncertainty, the axioms that are imposed are taken to be canonical of rational conduct. It is not, however, clear that this is the case. Indeed, it is relatively easy to construct counter-examples of reasonable agent behaviour which nevertheless contradict the neoclassical axioms of choice.

Take the following example. An agent at a party is offered the choice of one of several slices of cake of differing sizes. She likes the cake but does not want to appear too greedy either to herself or to the others present. So, in this case and in general, she adopts the following rule: choose the middle-sized slice. This rule is both sensible and reasonable to its ends (wanting to seem somewhat selfless) but it contradicts the strong axiom of revealed preference (assuming that the domain of choice is taken to be the range of slices of cake offered to her). Equivalently, the example contradicts the preference axioms.

Or consider the case of a regular churchgoer trying to determine how much money to give away to the charitable fund each week. The recipients of her charity are anonymous to her and she is ultimately unaware of exactly how the donation is to be spent; that is, the agent simply trusts thechurch to disburse the funds aptly to the ultimate recipients. The problem for the agent is to know how much to give—that is, she is confronted with the question: what constitutes 'charitable conduct' in this case? If it were the case that the average amounts donated each Sunday were posted at the back of the church the following week,

then the agent concerned to do the charitable thing might reasonably adopt the rule: donate this week an amount double last week's average donation. Again, this rule is sensible even though it contradicts any concept of choice or of preference over well defined objects.

It is worth noting here that, in these two examples, the problem is not one that has anything to do with the empirical stability of preferences but, rather, concerns what it is that agents have preferences over. Which is to say, the problem really goes to the issue of whether it is possible to have preferred ways of behaving which are not reducible to choosing the most preferred element from a set of things. This question clearly does not arise in neoclassical economics, since this takes commodities as primitive objects of theory (and therefore as being un-analysable) and takes it as axiomatic that the motivation for action is the desire to secure consequences over which the agent has consistent preferences (i.e. preferences which are complete, transitive and reflexive). Yet in these two examples, although the agents clearly have preferences for proceeding in one way rather than another, it is not at all clear what *things* they have preferences over. Indeed, as noted, in both examples the behaviour is not consistent with consequence-based motivation. The agents simply seem to have preferences for certain ways of going forward; i.e. agents seem to have preferences for rules of proceedure rather than preferences over well defined ends. Of course, in both cases, the desire to secure real consequences constitutes *part* of the motivational reasoning for proceeding in the preferred manner—the guest likes cake and the donor wants to help the poor in a material way—but a materialist-consequentialist account is inadequate to fully account for their actions, since in both cases the reason for acting also makes reference to some non-material objective: trying to be 'somewhat selfless' or 'relatively charitable'.

This presents preference-based theory with a problem, for the mooted behaviour is both sensible and reasonable, but it is unaccountable within a straightforward application of axiomatic reasoning (which is Joan Robinson's point). There are basically two ways of dealing with this reasonable deviation from the neoclassical axioms. The first is to abandon the (exclusively) consequentialist premises of neoclassical economics and replace them with (or at least supplement them with) a deontological motivation. On this account, while many actions are explicable in the terms of hypothetical—means-end—reasoning, of which neoclassical economics is a type, other behaviours are explained by reference to the categorical imperative. The problem with this solution is that in many cases, including those above, it is impossible to derive the specific rule used from the categoricalimperative. Which is to say, for example, that it is not at all clear why, in the first case, the guest ought to choose the *middle-sized* slice, or, in the second, why the donor should give *twice* the average donation, given just the premise: think oneself to be a universal legislator and act so that one could want all agents to act in like fashion. The

categorical imperative is sometimes too bombastic a principle to invoke in the explanation of only modestly ethical action. (In the first case, the guest only wants to be *somewhat* selfless, hence she does not eat the smallest slice; in the second, the donor wants to be *relatively* charitable but does not want to give away all that she owns. The problem for the deontological account is that it cannot explain this niceness of judgement.)[2] That being said, it does at least point us in the direction that is required; that is, it makes it clear that the materialist-consequentialist account lacks the capacity to convey the motivations and therefore the actual actions taken by agents in various situations.

The second method for dealing with reasonable 'deviant' behaviour is to redefine the nature of the objects over which agents 'really' have preferences (this is the essence of Hahn's remark). In the above cases, for example, degrees of 'selflessness' and 'charity' are attached to the states brought about by the agents' actions, and agents are assumed to have preferences over such objects as 'slices of cake and their associated degrees of selflessness' and 'donations and their degrees of charity'. Agents' behaviour is then held to be consistent with the axioms under this 'refined' description of objects. This approach greatly broadens the range of behaviours which are explicable in the terms of the axiomatic account.

This method of refinement of the description of consequences is, however, deeply problematical. To sceptical empiricist-minded neoclassical economists (e.g. Samuelson 1952:136; Kreps 1988:33 – 8), the ability to always find a description of objects for which the axioms hold is thought to be an empty victory, for then the theory can never be falsified. (The method of refinement takes any inconsistency with the facts, not as a partial refutation of the theory, but as a prompt to redescribe the objects of choice until the 'right' description is found.) Neoclassical economics then ceases to be a scientific theory at all—or at least, it ceases to be scientific in the Popperian sense, which is how neoclassical economists conceive their discipline.

There is a deeper problem than that which the empiricist sees, however. It is not merely an empty exercise to refine descriptions of objects, but an impossible one. For refinement to be possible, it is necessary that the new attributes (e.g. selflessness and charity) be self-evident properties of the state of affairs brought about by the agent(s). Which is to say, these properties need to be elementary—*a priori*—attributes of the objects of choice (as, for example, are the characteristics of commodities in Debreu's theory [1959], which are defined by their physical, temporal and locational attributes). Otherwise, rules of inference need to be given which permit the particular metaphysical/qualitative attributes to be inferred from the configuration ofthe choice situation. But to say that there are rules which allow agents—other agents or indeed the active agent herself—to read off from what has happened an appropriate and sensible description, is equivalent to saying that there is a grammar of action. (Thus, for example, we can tell that an agent is choosing to be somewhat selfless by not choosing the largest slice, or

that an alms-giver wishes to be relatively charitable by giving more than local norms demand.) The fact that agents' actions are to be understood as 'grammatical', however, has a number of radical implications for the manner in which we theorise about agency.

First, it implies that agents are essentially social animals in the following sense: their actions are intended to communicate intent. Regardless of whether the audience is constituted of other agents or simply consists of the agent regulating her own conduct, for communication to be possible there must be a set of publicly understandable rules—a social grammar—in which the communication is made. In Joan Robinson's terms, the intention that can be said to underlie agents' actions depends on the particular social situation and the associated 'rules of the game'. We understand agents' actions because they make sense within the context of the game and its rules. (Thus, for example, in the party and church 'games', we understand the agents' actions because we understand rules which guide their conduct.) Thus, *contra* neoclassical economics, it is not possible to take agents as defined independently of the society in which they are embedded. If agents' motivations are seen as being essentially expressive, then the grammatical rules of the game being played need to be known, along with the agents' roles in it, if we are to explain their behaviour. To put the matter pedagogically, in any foundational text on economic behaviour, the chapter on individual agency ought to come after—or in any case certainly not before—the chapter spelling out the social grammar/rules of the game. On this account, Robinson's view is fundamentally opposed to the reductionism of neoclassical economics (as, for example, spelt out by Hahn 1984:1 – 2).

The second point is that the rules of behaviour—the social grammar—are really *ways of proceeding* rather than rules for inferring appropriate descriptions of objects. Thus, for example, the grammar of selflessness implies that the guest does not choose one of the larger slices; and that of charity implies donating over and above the norm. The things that agents are after are not objects but ways of going forward. To put the matter in linguistic-grammatical terms, agents are after gerund-nouns, not object-nouns. For example, the guest wants to be seen as 'selfless' and the alms-giver, 'being charitable'. This itself points to an essential difference between the account of agency implicit in Joan Robinson's *Weltanschauung* and that of Frank Hahn. For the latter, each agent faces a pre-existing domain of choice over which she has well defined preferences, and her job is to choose the most preferred element. This account naturally lends itself to formalist techniques and takes its methodological lead from the hard sciences. According to the former view,on the other hand (i.e. Robinson's) 'actions speak intent' in the context of a well defined social grammar. This account is inherently constructivist in outlook and is more closely related to anthropology and linguistics than it is, say, to physics or biology. On this view of the world, there is not a pre-existing commodity space or quality space (or 'selflessness' space or 'charity' space) or any other kind of space over which agents have set preferences.[3]

Rather, agents have things they want to say, and their preferences for using one specific rule-of-action in a well defined context over another, simply represents their preference for saying one thing rather than another in that context.

Now, it is tempting for someone brought up on the axiomatic method to regard any alternative kind of analysis as being 'woolly' (or 'sociological'). But we have already seen a couple of examples in which this alternative method of analysis can be readily applied in the explanation of conduct which is otherwise unaccountable. Moreover, the modes of behaviour in these examples are readily reducible to formulae and so are capable of formal—though not axiomatic—analysis. Of course, it has to be admitted that these simple examples have a distinctly non-economic character. However, there are other social games in which the method outlined above can be applied and which are more obviously related to the field of economics. Take, for instance, the following two situations in which agents' behaviour is not directed to bringing about a state of affairs for its own sake but rather for the intent it communicates. (Or, to put the matter another way, each agent intends not to obtain a simple, *a priori*-defined end, but rather seeks to bring about a certain state of affairs, the significance of which depends on the social context in which it obtains.) The first case is the simple fictitious game: 'U, non-U'.[4] It has two players, A(ristocrat) and B(ourgeois) each of whom can play either right or left. It is the object of A to avoid being matched by B, and it is the object of B to match A. It can be played as a one-hit game or, more interestingly, as a supergame. There are no (simple) payoffs, nor are any necessary to explain the motion of the game (which, given the role of A, is one of non-determinate random motion). Yet this simple payoff-less game, and generalisations of it, can be, and have been, used to analyse a well known species of consumption behaviour (see Veblen 1899; Mitford 1956; Bourdieu 1984; Galbraith 1984).

Or consider the game 'novelty' in which agents demand goods purely because of their 'novelty value'. Again there is no set of entities over which preferences are defined. (Because newness is not an inherent property of an object there is no *a priori* space of objects with 'novelty properties' in which static preferences can be specified.) Yet, again, this game forms a well defined species of consumption behaviour (as every parent knows) and has been used to explain historical episodes of consumption (see Brewer and Porter 1993; Campbell 1987).

Both these examples are useful in understanding why advertising is, onthe whole, inconsistent with the axiomatic account, as asserted by Joan Robinson. For, in both examples, the behaviour of agents is rule- (not consequence-) directed, and it is relatively easy to see how certain kinds of advertising may reinforce the kinds of rules employed. Thus, for example, it is possible to conceive advertisements saying: 'If you want to avoid bourgeois habits of dress, wear this!' or, 'If you want to ape the aristocracy, buy this', 'This is the latest fashion!', etc. The point here is that advertising is either parasitic upon pre-existing rules of behaviour (which appear as kinds of imperatives) or it creates

new ones. Either way, the axiomatic account is inadequate to deal with this characteristic feature of advertising, for it cannot deal with imperative—i.e. rule-bound—behaviour in general.

The inability of neoclassical economics to deal with imperative behaviour follows from the fact that such behaviour is directed not to a predefined end, but aims at constructing propositions about agents' intentions in well defined social games. In the first example above, Aristocrats wish to signal that they are not bourgeois while the latter are keen to be seen in association with the former; while in the second, all agents intend themselves to be seen as stylish. Given the rules of the game these intentions can be signalled in well defined ways. But neoclassical economics cannot deal with this behaviour because it deals only in unanalysable ends and has no concept of a grammar of action. It bears repeating that this is a logical point, and it is well made by Robinson; but it is curiously misconstrued by Hahn. This is especially surprising since it is essentially an application of Wittgensteinian reasoning to economics.

The implications of the line of argument put forward in *Economic Philosophy* are not restricted to the purely analytical, however; there are ethical implications of which Joan Robinson was also well aware. As she noted, it is part of the scientific pretensions of neoclassical economics that it be thought of as ethically neutral:

> In all this kind of analysis, which is still taught and is still being elaborated with fresh embellishments, the notion of ethical judgement purports to be excluded and the whole exercise is put forward as a piece of pure logic. The very idea of moral implications is abhorrent to practitioners in this field.
>
> (Robinson 1962:57)

The argument that is used to defend the non-ethical character of neoclassical economics is essentially the one in which the description of consequences is 'refined' until just that set is found which allows the axioms to be applied. This method, as we have seen, allows any behaviour which seems to be imperative in nature—and so is *prima facie* inconsistent with the axiomatic approach—to be reinterpreted in terms of the optimal pursuit of desired ends. We have already seen this method applied to examples of selfless and charitable conduct, and it is easy to see how it applies to the other examples given above: aristocrats are taken to be maximising 'snobbishness', the bourgeois 'verisimilitude', and the shopper 'fashion', with utility being increasing functions of the defined variables.[5] This process of reifying ends makes all agents utility maximisers by definition, and therefore allows all action to be assessed purely in terms of efficiency. The theory is then able to maintain its silence on the ethical value of the ends themselves, while concentrating on the success of agents in achieving them. We have already seen that this operation is illicit since, in order to know what

attribute to reify, the observer must be able to 'read' the agent's actions—that is, the agent and the observer must share a social grammar. But this is not supplied by the theory, and indeed is inconsistent with an axiomatic account. Moreover, once the observer knows the grammar, then she is generally able to interpret the actions of the agent, because the particular behaviour corresponds to the application of a particular rule of action. The agent's actions and the interpretative rule are enough to explain the agent's conduct. Reifying—or hypostatising—the agent's objective adds nothing to our understanding of her behaviour. Thus, not only does reification rely on something absent from the theory (i.e. a social grammar) it is, in any case, redundant—that is, it falls foul of Ockham's razor. Which is to say that the whole exercise is meaningless since it adds nothing to our understanding of the agent's intentions and actions. It comes as no surprise, then, that neoclassical economics is silent about the ethical motivations of action: any theory which is vacuous can have nothing interesting to say.

To say that it is not logically possible for all agents' actions to be consequentially motivated is to deny Pareto optimality its primacy in welfare economics. For, once the concept of a preference map over consequences is found to be nugatory, the efficiency criterion simply vanishes as a concept. Joan Robinson saw this clearly and also saw what it implied: the reversion to a more explicitly ethical discussion of economic behaviour. In this—as in the more purely logical issues—she was evidently sympathetic to Wittgenstein's view, namely, in this regard, that the individual's ethical stance is not itself supported by reasoned argument but rather is the premise from which reason proceeds. As she has it:

> The upshot of the argument is that moral feelings are not derived from theology or from reason. They are a separate part of our equipment, like our ability to learn to talk…. Reason will not help. The ethical system implanted in each of us by our upbringing (even a rebel is influenced by what he rebels against) was not derived from any reasonable principles.
>
> (Robinson 1962:16 – 17)[6]

Thus our ethical discussion can never have the objectivity which neoclassical economics claims for the Pareto criterion. Instead, our assessments of any action remain rooted in the concepts and language of the society in which we have been raised. So, for example, our view on exactly what it is to be relatively selfless, somewhat charitable, snobbish, imitative, or fashion-conscious, and our judgements of the worth of these activities, is irremovably rooted in the conceptual framework of the society in which they take place. The point of normative analysis, then, is not to provide (falsely) objective advice on how to live or to organise a society; but is rather to provide the analytical tools necessary to allow the *interested* observer to weigh the pros and cons of any mooted set of

actions. Thus for economists the chief question of interest becomes not, what is the set of institutions which allows the given ends to be most efficiently achieved? But, what activities ought to be promoted and which curtailed given our ethical assessment of their worth?

This normative outlook on the role of economics is, needless to say, the opposite to that of modern mainstream economics. The latter—which sees economic activity as being principally about private individuals seeking private goods—is driven by its internal logic to a Paretian analysis of economic institutions: one institution is preferred to another according to how well it allows agents to pursue private ends. And, as every schoolboy knows, the institution which generally best solves this problem is the market. Yet if agents' actions are generally taken to be essentially *public* in nature—as Joan Robinson suggests—then the role of the market and of the state and of all the myriad other social institutions cannot be assessed purely in efficiency terms. Rather, the issue for these institutions is, what are the desirable kinds of social relations amongst the citizenry and between the citizenry and these social organisations (such as the state, the market economy, the education system, and so on)? In answering this set of questions, one must deal with the subsidiary issues: what institutions, old and new, allow the desired social relations to be best expressed? For example, should education be privately or publicly funded? (which cannot be answered without first answering the question, is education an essentially private good or a kind of acquired social skill?); should life prospects be equally or unequally distributed (say, in correlation with the distribution of wealth)? Should the unemployed receive income support payments (i.e. the dole) as a matter of right or should there be a quid pro quo (e.g. work-for-welfare schemes)? and so on. These questions are, strictly speaking, meaningless in terms of neoclassical economics since they presuppose a social grammar in which social relations are communicated (e.g. the social grammar of 'fairness', 'rightness', etc.). Yet, in ordinary public discussions, these issues are at the core of what interests the average citizen (or 'taxpayer', to use the modern, lopsided term). Of course, Joan Robinson's basic framework does not predispose one to any particular kind of answer to any of the given questions; but, unlike neoclassical economics, it does allow the relevant questions to be posed. Moreover, her analysis allows one to see why, at the most fundamental level, neoclassical economics is inadequate both as a 'positive' theory of choice and as a 'normative' guide to policy: it is too 'thin' to express the interests of actual economic agents.[7] The problem now is to develop a more broadly conceived model of agency which is both coherent in a way that the neoclassical model is not, and can provide a more humane foundation for policy formulation.

NOTES

1 Of course, when *Economic Philosophy* was published, in 1962, the Capital Debates were in their earliest stages and not all the issues had been fully worked out. Nevertheless, the point that the debates were concentrated upon logical and not empirical issues was already well known.
2 This is the basic conclusion of modern ethics; see for example Williams 1985; Taylor 1988.
3 Just as there is no such thing as a pre-existing 'space of sentences' nor preferences over such a domain, so there is no preference map of actions taken in a social context.
4 After Nancy Mitford's book.
5 This method of reification is particularly favoured in Becker's so-called 'New Economies' method—see for example his joint essay with Stigler (1977).
6 Compare these comments with those, for example, of Wittgenstein in *Culture and Value* and *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*.
7 The concept of the analytical thinness of the neoclassical theory of choice is Elster's (1983).

REFERENCES

Bourdieu, P. (1984) *Distinction,* London and New York: Routledge.
Brewer, I and Porter, R. (eds) (1993) *Consumption and the World of Goods,* London and New York: Routledge.
Campbell, C. (1987) *The Romantic Ethic and the Spirit of Modern Capitalism,* Oxford: Blackwell.
Elster, J. (1983) *Sour Grapes,* Cambridge: Cambridge University Press.
Galbraith, J. (1984) *The Affluent Society* 4th edn, Harmondsworth: Penguin.
Hahn, F. (1984) *Equilibrium and Macroeconomics,* Oxford: Blackwell.
Kreps, D. (1988) *Notes on the Theory of Choice,* London: Westview.
Mitford, N. (ed.) (1956) *Noblesse Oblige,* Harmondsworth: Penguin.
Robinson, J. (1962) *Economic Philosophy,* Harmondsworth: Penguin.
Samuelson, A. (1952) 'Utility, preference, and probability', in J.Stiglitz (ed.) (1966) *The Collected Scientific Papers of Paul A.Samuelson,* Cambridge MA: MIT Press, 127 – 36.
Stigler, G. and Becker, G. (1977) 'De gustibus non est disputandum', *American Economic Review,* 67, 2:76 – 90.
Taylor, C. (1988) *Sources of the Self,* Cambridge: Cambridge University Press.
Veblen, T. (1899) *The Theory of the Leisure Class,* London: Macmillan.
Williams, B. (1985) *Ethics and the Limits of Philosophy,* London: Fontana.
Wittgenstein, L. (1958) *Philosophical Investigations,* Oxford: Blackwell.

16
READING MARX
Joan Robinson's essay on Marxian economics

Prue Kerr

> the qualitative proportion means everything…. It comes to the surface here in a purely economic way…within the limitations of capitalist understanding, from the standpoint of capitalist production itself.
>
> (Marx 1975:259)

> Joan Robinson…always had a blind spot about what the labour theory of value really entailed.
>
> (Araujo and Harcourt 1995:112)

INTRODUCTION

In the Foreword to *An Essay on Marxian Economics* (1942), Joan Robinson points to 'the developments of economic life [which] have shattered the structure of orthodox doctrine…In my belief [orthodox economists] have much to learn from Marx' (1942:v). Shove (1944), in his review article on her *Essay* concludes, 'Mrs Robinson's commentary on Marx succeeds, as far as commentaries on this author do, in being sympathetic without becoming uncritical' (Shove 1944:50) but adds, by contrast, that in her 'onslaught on "orthodox economics"…she has allowed her moral sentiments to run away with her' (*ibid.:* 60). The present review suggests that neither was Marx spared her vitriolic judgement. But in making this judgement she errs on two major accounts, revealing her lack of understanding, and so leaving little of significance in Marx's analysis to be developed. First of all, her dismissal of the labour theory of value as 'awkward' and indeed 'obscure' creates superficial and often misdirected criticism in her account of Marx's capitalist dynamics. It also leads her to overlook the implications of the distinction between constant and variable capital. Furthermore, she identifies his theory of crises with his 'law' of the falling tendency of the rate ofprofits, and in subsequently rejecting this 'law' as a 'tautology' she rejects also his theory of crises. Her appraisal fails to understand the complexity of Marx's understanding of the critical features of

capitalism. The *Essay* presents a view of Marx which pervaded her later ideas and which is still prevalent.

From the outset her motivation is suspect, the investigation selective. She begins humbly,

> I have confined myself to Marx's economic analysis in the narrow sense, and made no attempt to deal with the broad treatment of history and sociology which form the most important part of Marx's doctrine…it is true that no particular aspect of Marx's argument can be properly understood without a grasp of the whole.
>
> (Robinson 1942:6)

The preface confines her scope to what must be necessarily a limited exposition of Marx's theory of capitalist dynamics. However, a brief diversion into her use of language gives a hint as to the sympathies of her immanent argument. Marx's 'unscientific' mode is confronted throughout. In the preface, she states 'The chief difficulty learning from [Marx] arises from the peculiar language and the crabbed method of argument…and my purpose is to explain what I understand Marx to have been saying in language intelligible to the academic economist' (1942:vi). Yet throughout the book her own choice of language resurfaces to replace argument where she fails to understand what Marx is doing. Consider just a few examples: she describes the 'simple dogmatism of the first volume of *Capital*' and 'the purely dogmatic statement' of the labour theory of value as a basis for understanding relative prices. Accepting Marx's perspicacity but not his analysis, she pronounces 'The substance of Marx's argument is…[not] irrelevant to the modern situation, but the argument has become incompatible with its verbal integument' (1942:22), and, 'His terminology derives its force from the moral indignation with which it is saturated' (*ibid:* 26 – 7). Robinson is vehement in her rejection of the labour theory of value and in the possibility of existence of a non-falling rate of profit. Her rejection of the labour theory of value refers to the 'trouble [which] probably arose, like most of the obscurities in Marx's argument, from his method of reckoning in value' (*ibid:* 46). On the second issue, Robinson finds 'confused and redundant' Marx's tendency for the long-period rate of profits to fall (*ibid:* 6, 26 – 7). The confidence of her rhetoric presumes a monopoly over accuracy, over 'scientificity'.

MARX'S SOCIOECONOMIC STRUCTURE

Marx begins from general premises which are true of all modes of production; he then examines the stages of society preceding capitalism and asks what it is that makes capitalism different from these. One critical factor is the monopoly by some over the ownership of capital, the means of production and of finance, and the necessary sale by those without capital, of their labour power, as a

commodity for exchange. The labour theory of value describes this production relation as it evolves from an exchange economy of simple barter to a capitalist market. It underlies Marx's social relations of production and exchange. Shove (1944) criticised her mistaken attempt to extrapolate too literally Marx's labour theory of value in Volume I of *Capital* to his discussion of relative prices in Volume III.

A number of social/economic relationships particularise capitalism. These include the relationships of antagonism between capital and labour, of competition between capitals and of interdependence and collusion between capitals. Historical phases, characterised by various political and economic circumstances, are differentiated by the manifestation of these relationships. Attempts to resolve their contradictory co-existence gives rise to new, ambivalently appropriate forms within each phase. Marx's method of analysis thereby traces out the stages of development of labour through its material history as labour moves towards its commodity form under capitalism. Basic to his explanation of this present phase and its implications for capitalism, is the labour theory of value.

But Robinson dismisses such a proposition as 'mumbo jumbo' and so rejects this theory as a starting point. At the same time, Marx's construction of general, abstract concepts and their resolution and evolution to specificities of capitalism, are seen as unnecessary. She, like Marx, goes straight to the observed, but her subsequent abstractions miss the structural forces imperative to Marx. Robinson describes the historical evolution of society to capitalism through the Classical successive stages but without the insights derived through Marx's social relations of production and exchange. Ultimately, without a labour theory of value, she fails to appreciate the implications of the generalisation of commodity exchange to the market for labour power. Her definition of the alienated surplus is therefore a general one of an 'ever-increasing difference between total output and total real wages' (1942:11). This indeed dismisses any historical or political determinism of wages or the surplus. Yet it takes capitalism to ensure the surplus is appropriated from its producers in the form of value, and accumulated and used to purchase further labour power; profits have first to be produced—they are not simply a distribution outcome.[1]

Similarly, she concludes that Thus the formulation of Volume I slurs over a number of problems which are clearly distinguished in Volume III' (1942:17). Even more strongly, 'As I see it, the conflict between Volume Iand Volume III is a conflict between mysticism and common sense' (*ibid.:* 18). With these words she dismisses Marx's Hegelian logic. Had she acknowledged Marx's method, her critique of orthodoxy's search for a 'general theory' could have been better grounded and more devastating to them.

THE LABOUR THEORY OF VALUE

The labour theory of value, her primary target, is therefore irrelevant to her. To demonstrate its pedagogical damage is one of her fundamental objectives:

> [I wish to show] that no point of substance in Marx's argument depends on the labour theory of value…the labour theory provides the incantations.
>
> (Robinson 1948:27)

Her view of the labour theory of value is, as an 'academic economist', that it is the theory of relative prices. As such, it is neither

> particularly important [n]or…fundamentally opposed to orthodoxy.
>
> (*ibid.:* 52)

According to Robinson, Marx ultimately abandons his alleged view that 'prices correspond to values under capitalism' (1942:27). As a consequence of this misconception much of her argument is irrelevant. Her exposition of Marx's labour theory of value begins with a quote from Marx pointing out that the contingency for exchange is use value—a 'useful article'—and this article has value only because it embodies human abstract labour. This is a statement about labour and production and exchange in general, not one particular to capitalism. Marx then asks 'how is the magnitude of this value to be measured? Plainly, by the quantity of labour contained in the article' (cited in Robinson 1942:14). Robinson's answer to this is simple at this level of abstraction. The value of a commodity consists of the labour-time required to produce it, including the labour-time required by subsidiary commodities which enter into its production. Indeed, if it is just that, this view demonstrates her own failure to recognise Marx's method of moving from general production to the particularities of capitalism. It signifies that labour is the only source of value; that the value of raw materials comes from the labour power required to render them useful for production and not from nature; and neither does value come from owning assets. This is the most general statement of Marx's labour theory of value.

The labour theory of value, therefore, underlies the explanation of societyas it evolves through successive stages which precede capitalism to the various institutional forms which capitalism itself takes, and thus the origins of profits and indeed the rate of profits itself. But Robinson sees this theory as definitional.

She then refers to Volume I of *Capital* to guide her on the relationship of value, an abstract concept, to price, a material concept appropriate to capitalism. Her failure to understand the different referential points of the two concepts is apparent in her statement, 'The definition of value has to be stretched and strained a good deal in order for Marx to maintain that prices tend to correspond to values' (1942:16).

She pre-empts her own interpretation of Marx's argument with the following propositions; This ratio [s/v, the rate of exploitation] plays the leading part in Marx's whole argument…[it] is unambiguous' (1942:8). Its primary importance to her is that it indicates distributive shares. But such symbols are one-dimensional. While she observes that it is surplus labour time which underlies the amount of the working day devoted to surplus production, the (economic) implications of the historical/political specificity of wage labour, appropriation of the surplus as surplus value and, ultimately, the different levels of antagonism between the interests of labour and of capital are overlooked, and so, by implication, must be the dynamic nature of the capitalist process.

The rate of exploitation, expressed in the ratio *s/v,* is the first of three ratios by which she appropriately characterises Marx's theory. The second is c/v, the organic composition of capital, and the third is *s/(c+v),* the rate of profits. All concepts are specified as rates per unit of time. (Note that c represents depreciation and raw materials rather than the stock of capital invested so that p′ is the share of profits in turnover, a flow concept. Marx was aware of the distinction as he sometimes wrote s/(c+v) to refer to the rate of profits on capital invested.)

The ratio, capital per labourer—*(c+v)/v*—can vary in three ways:

1 In a slump, employment falls while the fixed capital stock remains constant (except for depreciation);
2 accumulation tends to increase capital per labourer for given degrees of capacity utilisation; and
3 technical progress, changes in the rate of interest and in the wage rate could each make capital per labourer vary in either direction.

Hence the rate of profits might fall or rise, although the stronger tendency is likely to make it fall. None of these observations is controversial. Marx discusses the variations of these ratios and their interactions throughout Volume III of *Capital*. His conclusion is of counteracting effects and so tendencies and, in the short run at least, inconclusive outcomes. He confronts the possible permutations. Robinson sees this inconclusiveness asproblematic. For Marx it illustrates the contradictory dynamic of capitalism ultimately explicable in terms of the labour theory of value. The complexity of capitalism makes knowledge of precise outcomes impossible. Robinson is committing that most common error, which is to refute Marx on the basis of the failure of his alleged predictions; most importantly for her, the failure of the rate of profits to fall. But Marx had theories about possible tendencies, not predictive inevitabilities. Yet, ironically, Robinson shares this shadowed vision.

Her misgivings about Marx, specified thus, are that he

1 ignores the impact of capacity utilisation (‘he assumes that the capacity output of a given amount of capital is rigidly determined by technical conditions’ (1942:10), and
2 he assumes that the rate of interest has no effect on the capital structure, and on the wage rate ‘only indirectly’ *(ibid.),* and this through its influence on technical progress.[2]

With regard to the first objection, Marx assumes a tendency to a given capacity-output ratio in the ‘long period’, but in the ‘short period’, firms are operating at different actual rates of profits. Smaller and less profitable firms are losing markets and operating at lower capacity, or they may be operating old machinery at high capacity and high cost, aiming to amass profits to survive rather than to maximise a profit rate. This is Marx’s process of the dynamics of concentration and the counter-tendency to a falling rate of profits.[3]

Consider her second objection, to Marx’s neglect of the role of the interest rate. It is not through relative prices of labour and capital that the new ‘marginal productivities’ associated with technical progress lead to capital substitution. Marx is aware of the politics of ‘full’ employment, for example, for the ‘discipline enforced by the capitalist for the combined labour’ (Marx 1975:83). Furthermore, stagnation means to lay off workers, while those remaining employed ‘would have to submit to a reduction of wages even below the average’ (*ibid.:* 254). Stagnation, however, is hardly conducive to increasing the fixed capital component of production through new investment. It might revive a falling rate of profits by encouraging mergers and take-overs. And most obviously there are the contradictory possibilities with regard to producing a surplus, of substituting machines for wage labour and, at the same time, searching out new labour-intensive industries to restore the declining surplus value.

Her lack of insight into Marx’s method of proceeding is again summed up in her remark ‘There is no reason why the rate of exploitation should be treated as either logically or historically prior to the rate of profits’ (Robinson 1942:19). This contrasts with Marx’s ‘values are not only theoretically but also historically prior to price of production’ (Marx 1975:177).Marx describes the relationship between value and market price so that market price is governed, but not at any moment precisely determined by, the labour theory of value. First of all,

> The exchange of commodities at their values, or approximately at their values, thus requires a much lower stage [of social progress] than their exchange at their prices of production, which requires a definite level of capitalist production.
>
> *(ibid.)*

But Marx is specific about the role of the labour theory of value for exchange values, and states that to say that commodities

> are sold at their values merely implies…that their value is the centre of gravity around which their prices fluctuate.
>
> (*ibid.:* 178)

The price of production integrates the average rate of profits on production with the socially necessary labour cost of production. The competitive tendency to the equalisation of rates of profits implicit in this concept suggests a particular state of development of capitalism. Final market price is approximately the same price paid for the same commodities, no matter what their relative conditions of production are or the quantities supplied and demanded are. Where market supply and demand coincide, the market-or short-period price equals the price of production and the formation of prices from values follows the rules of capitalism. And thus Marx develops his abstractions with reference to gradually emerging capitalism with labour values underlying prices of production, 'the centre of fluctuation for market prices' *(ibid.)*.[4]

THE FALLING RATE OF PROFITS

Robinson also targets Marx's theory of crises. Marx's theory, she argues, employs three elements contributing to crises: a theory of the reserve army of labour, whereby as the stock of capital accumulates for a given labour supply, the level of employment falls; a theory of the falling rate of profits, according to which as capital accumulates the average rate of profits falls; and the theory of overproduction of capital relative to wage goods and the incapacity of too-low-paid workers to purchase these goods with their wages. While Robinson sees 'some resemblance' between the first and third factor, she isolates the second as both fundamental to Marx's theory and questionable. She identifies it with his theory that there is a tendency for the rate of profits to fall. Compounding this problem, 'Marx's argument fails toestablish a presumption that the rate of profit tends to fall, when the problem of effective demand is left out of the account' (1942:47).

Consider first the relation between wages and the rate of profits. She objects that if real wages remain constant, p' rises or falls, as (c+v)/v increases accordingly as the ratio of proportionate increase in product to the proportionate increase in capital exceeds or falls short of the ratio of profits to product. Again she misses the significance of the respective concepts of constant and variable capital, and then of price of production and market value, that Marx develops to construct his argument illustrating the ultimate relation between labour value and final price. The rate of surplus value may be uneven between industries but the rate of profits will tend over the long run to be equalised by the movement of

capital. The impact of this tendency depends, among other things, on the sophistication of finance capital—a point of which Marx was aware.

With the development of capitalism, the rate of profits will tend to fall as more and more constant capital is used to work with given labour and the potential for producing surplus value decreases. Robinson explains this simple corollary of the labour theory of value in terms of 'diminishing returns', which fortunately she dismisses as not compatible with Marx's theory about technical progress. Robinson argues that if knowledge develops as capital accumulates, there need be no tendency for diminishing returns, and with constant returns there can be no tendency for the rate of profits to fall (assuming no problem of effective demand and the short-period rate of profits). The most that can be said, she argues, is that periods of falling profits may occur when capital per labourer increases very rapidly relatively to the rate of advance in technical knowledge. In Marx's view, however, technical knowledge is not an 'exogenous' factor, and, when accumulation is rapid and labour becomes scarce and politically stronger, a strong stimulus to labour-saving invention develops. Such an increase in the proportion of fixed or constant capital to wage labour might reduce the rates of surplus value and of profits while increasing the mass. But again, Robinson ignores Marx's explanation of the production of surplus value. Again she misses the significance of the labour theory of value in differentiating the concepts of constant and variable capital. So she can state

> Moreover, the whole apparatus of the theory of value is designed to exclude the notion of attributing productivity to capital, and allows no room for the concept of the marginal productivity of a particular factor. A theory of falling profits based on the falling marginal productivity of capital would be something quite different from Marx's theory.
>
> (1942:45)[5]

Quite so!

On the other hand, for orthodox theory, a falling tendency in profits entails a rising tendency in wages. For the orthodox economist this presents no difficulty, but for Marx it is a stumbling block (Robinson 1942:43). It is not clear whether she is arguing that orthodoxy and Euler's theorem have more to say than Marx, so juxtaposing conceptually incongruous logical structures. But since Marx is isolating tendencies rather than laying down laws, he often invokes *ceteris paribus* while he explores changes in other ratios, for example in the concentration of capital. Despite short-period frictions which are barriers to equalisation of the rates of exploitation, he argues, the long-period tendency in capitalism is to equality. Furthermore, there are limits to the offsetting tendencies and these are eventually insufficient to offset the falling tendency of the rate of profits. Marx states

> the compensation of the reduction in the number of labourers by means of an intensification of exploitation has certain impassable limits. It may, for this reason, check the fall of the rate of profit, but cannot prevent it entirely.
>
> (Marx 1975:290)

Robinson, to the contrary, responds with these words:

> the rise in the rate of exploitation…through a rise in productivity, with constant hours and intensity of work, and constant real wages, is not limited in the same way. Productivity may rise without limit, and, if real wages are constant, the rate of exploitation rises with it. Marx appears to have been in some confusion upon this point, for when he begins to discuss the effect of a rise in productivity on the rate of exploitation, he switches… the argument to discussing the effect of changing the length of the working day.
>
> (Robinson 1942:46n)

But Marx, in discussing technological change, argues that if the new technology reduces the value of constant capital, it cheapens the elements of production (either those directly involved in production or those which comprise wage-goods), and then the same capital value can produce an expanded mass of commodities. The contradiction here is that although the depreciation of existing capitals brings with it a corresponding expansion in the mass of capital values, the new technology introduced can embody a higher rate of profits while also tending to embody a higher composition of capital. As the new methods reduce socially necessary labour time, the least efficient firms will drop out as more efficient firms hold their market prices down to squeeze out those competitors lacking their own advantage in technique. So, Marx 'answers' Robinson's objections to the compatibility of technical change with a falling rate of profits:

> No capitalist ever voluntarily introduces a new method of production, no matter how much more productive it may be, and how much it may increase the rate of surplus-value, so long as it reduces the rate of profit.
>
> Yet every such new method of production cheapens the commodities. Hence, the capitalist sells them originally above their prices of production, or, perhaps, above their value…. His method of production stands above the social average. But competition makes it general and subject to the general law. There follows a fall in the rate of profit—perhaps first in this sphere of production, and eventually it achieves a balance with the rest—which is, therefore, wholly independent of the will of the capitalist.
>
> (Marx 1975:264 – 5)

Marx argues that capitalists must introduce the newest technology in order to remain competitive or to try and temporarily gain advantage by selling at less than the (average) socially necessary labour time. Robinson mistakenly concludes that the rate of profits must always be increasing with new techniques. But so long as less living or direct labour is involved in the new methods, there will be less potential for the production of surplus value in that industry. She has again failed to see beyond the mechanical relation between the mass and rate of surplus value on the one hand, and between constant and circulating capital on the other. Furthermore, she has isolated the tendency for the rate of profits to fall as the singularly necessary cause of crises, thereby enabling her to dismiss Marx's theories of crises together with her dismissal of the labour theory of value and the falling rate of profits.

EFFECTIVE DEMAND

Robinson's alternative explanation of crises lies in effective demand. Marx's Classical schema of reproduction illustrate the mechanisms whereby a level of effective demand leads to a certain level and composition of output and employment, and to the distribution of this output between wages and profits and between investment and wage-goods. These relationships in themselves do not constitute a theory of output, but a framework which shows the essential interdependency between capitalists, capitalists and workers and the importance of accumulation for the survival of capitalists as a class. Marx, for example, uses the notion of effective demand, in a Keynesian sense, in the following passages, referring to 'social demand (this always taken to mean the effective demand)' (Marx 1975:181). It is

> essentially subject to the mutual relationship of the different classes and their respective economic position, notably therefore to, firstly,the ratio of total surplus value to wages, and, secondly, to the relation of the various parts into which surplus value is split up…. And this…shows how absolutely nothing can be explained by the relation of supply to demand before ascertaining the basis on which this relation rests.
>
> (*ibid.:* 181 – 2)

The problem of a lack of effective demand arises 'because in this specific capitalist interrelation the surplus-product assumes a form in which its owner cannot offer it for consumption, unless it first reconverts itself into capital for him' (*ibid.:* 257). And that there is a problem of realisation of surplus-value which can threaten the process of expanded reproduction is clearly stated:

> this production of surplus-value completes but the first act of the capitalist process of production—the direct production process…. Now comes the

> second act of the process. The entire mass of commodities, i.e., the total product *(c+v+s)*…must be sold.
>
> (*ibid.:* 244)

These were not particularly profound observations, as earlier writers had made similar remarks. But Marx, at an abstract level, suggests a reason for the emergence of this problem of inappropriate or insufficient effective demand for enabling uninterrupted accumulation to continue.

> The conditions of direct exploitation, and those of realising it, are not identical. They diverge not only in place and time, but also logically. The first are only limited by the productive power of society, the latter by the proportional relation of the various branches of production and the consumer power of society. But this last-named is not determined either by the absolute productive power, or by the absolute consumer power, but by the consumer power based on antagonistic conditions of distribution…. It is furthermore restricted by the tendency to accumulate, the drive to expand capital and produce surplus-value on an extended scale.
>
> (Marx 1975:244)

This consumer power is itself at variance with the narrow basis on which the conditions of consumption rest, that is, a sufficiently large wage share. In addition, the failure of investment and accumulation more generally compounds this tenuous nature of the connection between production and sale. The other side, then, is that as capitalist production develops, it 'intensif[ies] the contradiction between the conditions under which this surplus-value is produced and those under which it is realised' (*ibid.:* 245).The contradiction to which Marx refers here is that resulting from the tendency for the organic composition of capital to rise with the development of the productive forces, so that out of the surplus an increasing proportion is allocated to constant rather than variable capital, while at the same time it is from the latter that further surplus-value is produced; the wage-goods industry is shrinking from both production and distribution forces.

Although a certain amount of surplus-value might be produced, the resultant accumulation will depend on what portion of it is reconverted to capital, and out of this, to variable capital. This latter will be affected by the further introduction of new technology. The new technology tends to involve a higher ratio of constant to variable capital, of machines to labour. But it is the total wage bill, the variable capital, which provides the means for purchasing the products of labour. Whether this is greater or less in real terms than before will depend on whether the technological change cheapened wage goods proportionately more than it reduced total employment, and whether there is for some reason saving

from the surplus in either sector. Yet Robinson interprets Marx as confusing the effects of two of his theories:

> his theory of the falling tendency of profit, based on the principle of the rising organic composition of capital…is inextricably mixed up with the underconsumption theory, and the two lines of thought are not brought into any clear relation with each other. The theory of the falling rate of profit is a red herring.
>
> (Robinson 1942:59 – 60)

It prevented Marx from properly developing 'Keynes'' theory of effective demand. But the 'inconsistencies' she points out are further examples of his assuming *ceteris paribus* to explore possibilities. He argued that if growth is accompanied by uniform accumulation of capital and a given relationship between productive capacity and the stock of capital goods, there is no problem of effective demand. Robinson recognises elements of a theory of effective demand in Marx's use of the schema, in Volume II of *Capital*. She correctly sets out his argument as follows: if part of the surplus of both Group I (investment goods) and Group II (wage goods) is saved, that is, not spent on consumption goods, then V_1+S_1 exceeds C_{II}, and must be matched by an equivalent outlay on new capital goods out of S_{II}: 'Saving represents sales without purchases, and can proceed smoothly if it is offset by equivalent investment—purchases without sales' (Robinson 1942:56). Such a balance, Marx argues, 'is an accident under the crude conditions of [capitalist] production' (cited *ibid.*). Over-production of capital is never anything more than over-production of means of production—of means of labour and necessities of life—which may serve as capital, i.e. may serve to exploit labour at a given degree of exploitation (see Marx 1975: Volume III, 255 – 6). Robinson, on the contrary, sees the problem of effective demand asa problem of 'the distribution of income, between wages and surplus…such as to set up a chronic tendency for a lack of balance between the two groups of industries [consumption goods and investment goods]' (1942:58).

However, in such a way, for Marx, 'the entire nature of the capitalist mode of production is lost sight of; and also forgotten is the fact that it is a matter of expanding the value of capital, not consuming it' (Marx 1975: 257). Accumulation, or expanded reproduction, is a crucial aspect of the competition between capitalists and of the survival of capital as a whole; to maintain a certain mass of profits in the face of a falling rate of profits (in Marx's argument) requires constantly expanding reproduction. The excess accumulation of capital has as its corollary the production of excess commodities, which may have use value to the mass of unemployed but have no exchange value.

THEORY OF INVESTMENT

Robinson's failure to appreciate these differences between Keynes and Marx leads her to criticise Marx for not having a theory of investment or, at best, a theory which related investment to the (impossibly falling) rate of profits. Nevertheless, when she examines orthodox theory she is more generous with Marx (perhaps because she has to draw a similar conclusion about Keynes):

> The lack of a clear treatment of the inducement to invest is…a weakness in his treatment of crises, but from a long-period point of view it may well be that it is unimportant, and that any prospective level of profit, within very wide limits, is sufficient to keep the system running. Mr Keynes puts forward…[the same view as Marx]: 'Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of animal spirits—of a spontaneous urge to action…and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities'.
>
> (Robinson 1942:73, quoting from Keynes' *General Theory*)

Marx argues the imperative for capitalism to accumulate. And, Robinson observes, 'Mr Kalecki's epigram "The tragedy of investment is that it causes crises because it is useful" has a close affinity with Marx: "The real barrier of capitalist production is capital itself"'(Robinson 1941:235). Whether these comparisons do Marx's theories justice is questionable but Robinson does see in his use of the schema the suggestion that Marx was on the track of the idea that variations in investment are the key to the trade cycle. 'Marx does not develop a full theory of the trade cycle, or of the LR movement ofcapitalism, but he points the direction in which a theory can be found' (Robinson 1942:56). A year earlier she had written:

> Volume II [of *Capital*] also contains a detailed analysis of the way in which the process of investment generates purchases without sales and so promotes boom conditions…and a hint…that the length of the cycle may be connected with the average length of life of plant. All this suggests that Marx was on the trail of the analysis of effective demand. But instead of following it up he turns, in Volume III, to a fresh scent—the law of the falling tendency of the rate of profit.
>
> (Robinson 1941:240)

But what she still fails to see is that Marx demonstrated how stagnation, how disruptions in the smooth process of expanded reproduction, 'trade cycles', and technological change, each a contributor to crises in its respective way, followed from the logic of the economic structure, that the resolution of opposing forces was

itself problematic: that 'dynamic' is not the simple dichotomous alternative to orthodoxy's 'static'.

CONCLUSION

It is interesting to see the slightly more cautious 'Introduction' to Luxemburg's *Accumulation of Capital* that Robinson wrote. There, she stated that her purpose was to

> search for the main thread of the argument…and set it out in simple language. The result is no doubt too simple. The reader must sample for himself the rich confusion in which the central core of analysis is embedded, and must judge for himself whether the core has been mishandled in the process of digging it out.
>
> (Robinson 1951a:13)

Perhaps the impetuosity of the *Essay* became tempered by new insights. In 1951 she also wrote that one view of the labour theory of value had its place in Volume I of *Capital* where Marx assumed uniformity of c/v in all industries. 'Marx was concerned [with] movements in the total supply of capital and labour, the development of the productivity in the economy as a whole, and the distribution of the product of industry between labour and capital' (Robinson 1951b:147). However, she later rejected this view: 'It was not the way Marx looked at the matter' *(ibid.)*. In fact, to the end of her life she depicted the labour theory of value as 'the slogan that only labourproduces value' (1979:248 – 9). Crises were contingent on inadequately analysed effective demand.

Her final writings on Marx were much inspired by Sraffa. She was to see Marxists as 'losing sight of the main point', distracted by the intricacies of the transformation problem when they could be developing their own critique of the orthodox theory of supply and demand as distinct from a Marxian theory of the price level. Sraffa, she argued,

> does not attempt to offer a theory of the rate of profit or the rate of exploitation. Sraffa offered his argument as a battering ram to knock down orthodox theory and clear a space in which the Marxian theory of distribution and of prices could be elaborated.
>
> (Robinson 1979:250)

Robinson's *Essay* is mostly about Marx but it also takes the opportunity to criticise orthodoxy with reference to why Marx's approach (if not his theory) is superior. For example, one dimension of her critique, which was to become a recurring theme for her, is the lack of reality in the orthodox arguments, their preoccupation being with the 'search for eternal principles…[and]…elaborations

of minor problems' *(ibid.)*. Another was their confusion between comparisons and changes over time. Marx, to the contrary, she points out, argues from premises which are based on historical and contemporary features of capitalism; his specification of major problems for capitalist economies is the superior as his theories explain actual processes. Perhaps her institutional reading of Marx extended to her own positive contributions to areas of theory in development economics, growth theory and methodology.

Reluctant to abandon Marx, she says

> With the light that Sraffa has thrown on the theory of value and Kalecki on the process of realisation of the surplus we can develop a complete system, not of neo Marxism but of intelligible Marxism and…adapt it to the analysis of contemporary problems of capitalism.
>
> (1979:253)

Keynes, of course, is ever pervasive, not only compatible with Marx but superior. Keynes is the one who solves the Classical problems of accumulation, effective demand, investment and the determination of the rate of profits. In her allusion to Marx's vision of 'gloomy grandeur', she misconstrues his political convictions and optimism for a better society after the collapse of capitalism, reflecting her own (and Keynes') sympathies for the possibility of a humane capitalism.

Robinson may have been more open to Marx's arguments if she had notto first extract the 'economies' from the already narrow scope determined by the discipline from which she comes. What this does is prevent her from appreciating Marx's 'dynamics' and hence the implications of much of what he argues when indeed her true sympathies lay within his 'economic' embrace.

NOTES

1 'Basically, she insists on seeing the labour theory of value as a theory of relative prices rather than as a portmanteau term for Marx's explanation of the origin of profits in capitalism' (Araujo and Harcourt 1995:113n).
2 Robinson is possibly reflecting remnants of her own, still orthodox, ways of thinking.
3 In Marx's words, 'to increase the productivity of a given quantity of labour, to lower the proportion of variable to constant capital, and thereby to release some labourers…would tend to raise the rate of profit…The ensuing stagnation of production would have prepared—within capitalistic limits—a subsequent expansion of capital' (Marx 1975:255).
4 Cf. Araujo and Harcourt 1995:112.
5 Her own ongoing struggle between orthodoxy and concern with socially and economically evolutionary processes is revealed in her observation that

> the equilibrium rate of profits is that rate which induces zero net investment. But over the course of history since the Industrial Revolution began, net investment has always been going on. The actual rate of profits, therefore, good years with bad, has exceeded the equilibrium rate. Abnormal profits are the normal rule.
>
> (Robinson 1942:60 – 1)

See Araujo and Harcourt (1995), for an examination of the letters between Dobb, Shove and Robinson discussing this provocative statement and their own reformulation of it in terms of Robinson's subsequent model of growth and distribution.

REFERENCES

Araujo, J. and Harcourt, G.C. (1995) 'Accumulation and the rate of profits: reflections on the issues raised in the correspondence between Maurice Dobb, Joan Robinson and Gerald Shove', in G.C.Harcourt, *Capitalism, Socialism and Post-Keynesianism: Selected Essays of G.C.Harcourt,* Aldershot: Edward Elgar, 79 – 106.

Marx, K. (1975) *Capital,* vol. III, London: Lawrence and Wishart.

Robinson, J. (1936) 'Some reflections on Marxist economics. Review of John Strachey's The Nature of Capitalist Crises (1935)', *Economic Journal,* vol. 46: 298 – 302.

——(1941) 'Marx on unemployment', *Economic Journal,* vol. LI: 234 – 48.

——(1942) *An Essay on Marxian Economics,* London: Macmillan.

——(1948) 'Marx and Keynes', *Critica Economica,* November. Reprinted in Robinson (1974) *Selected Economic Writings,* Bombay: Oxford University Press.

——(1951a) introduction to R.Luxemburg, *The Accumulation of Capital,* London: Routledge and Kegan Paul.

——(1951b) 'The labour theory of value', in Robinson (1975) *Collected Economic Papers,* vol. 2, Oxford: Blackwell, 49 – 58.

——(1979) *Collected Economic Papers,* vol. 5, Oxford: Blackwell.

Shove, G. (1944) 'Mrs Robinson on Marxian economies', *Economic Journal,* vol. XLIV: 47 – 61.

17
JOAN ROBINSON
A neglected precursor of internal migration models

Pervez Tahir[1]

INTRODUCTION

The title of this essay succinctly states its purpose: to show that the work done by Joan Robinson on labour mobility in the thirties is an important precursor of the internal migration modelling started in the late 1960s to explain the rising urban unemployment in the developing countries. It is pointed out that the Hicksian wage differential approach had been challenged by the chance-to-be-employed approach long before Harris and Todaro attempted to substitute their employment probability model for the Lewisian wage gap model.

The next section outlines the commonly known roots of internal migration models. It will be seen that while economists have neglected Robinson's work, geographers and demographers do mention it. A selection of the statements of their representative positions shows that, so far as the search for an alternative to the wage differential as the principal determinant of internal migration is concerned, Robinson, T.W.Schultz, and Harris and Todaro are members of the same set; the fact that Robinson was not specifically concerned with a developmental context is of no consequence. The third section examines the Robinson hypothesis in some detail and in juxta-position with the established Todaro hypothesis of the determinants of internal migration. The incentive to migrate, in the Joan Robinson world, is provided by the chance to be employed, irrespective of the relative wage differential. The story begins, like all Keynesian stories, with an increase in investment that in this case is at a destination; it increases employment and thus attracts workers from an origin with relatively higher unemployment. There are two sub-plots, one relating to the case of an unchanged money wage and the other to a rise in it. After suggesting the determining influence of the chance to be employed in these cases, the section turns to the differences in the conceptualisation of this chance. It is relative unemployment in the Joan Robinson story, as opposed to the destination rate of employment in the Todaro hypothesis as well as the Harris-Todaro model.

Towards the end, the section also takes a comparative view of some post-Todaro approaches and the Robinson view.

In the fourth section, the essay looks at the consequences of migration. It is noted that the effects of migration have received far less attention than the determinants in theoretical as well as empirical research. In this context it is significant that Robinson had suggested two important effects: first, migration itself influences employment, and second, the initial differences between origin and destination may become cumulative. The fifth section attempts to reconstruct Robinson's early migration story in the light of her later work. Although she herself seems to have neglected migration as a topic of major interest in her subsequent writings, the reconstruction has been possible nonetheless for two reasons. First, her view of the price implication of increasing employment, on which the earlier migration story was based, never changed. Second, several scattered pieces of later writings on development either contain relevant insights or have a bearing on the subject of internal migration.

THE KNOWN ROOTS OF MIGRATION LITERATURE

Todaro complained about the dominance of geographers, demographers and sociologists in the field of research on migration (1976:25). The irony is that economists, Todaro included, have ignored Robinson's attempt to deal with the question of why work people move, while geographers (Parr 1966: 155) as well as demographers (Suzuki 1970:338) do take note of her work. As for sociologists, Todaro's black-out on the early essential work of Owen (1937) is as complete as on Robinson's contribution. Tracing the lineage of his model (Todaro 1971:393 – 4), Todaro refers to the original formulation by Hicks (1932: 76) of the real wage differential approach and its use by Lewis (1954) in the case of developing countries. His contention is that 'such an analysis is not very realistic in the context of the institutional and economic framework of most Third World nations' (Todaro 1976:29).

The important contextual difference is caused by the assumption about employment. Assuming full employment in the Lewis case means that a worker deciding to migrate from the traditional to the modern sector moves straight into a job. Thus the decision to migrate is not a function of the quantity of employment but of the relatively higher price of labour in the modern sector. Todaro maintains that rising urban unemployment belies the full employment assumption, so that the wage differential is no longer the sole determinant of migration. The mechanism of adjustment is not price, but some quantity.

In his search for an historical illustration of quantity adjustment, Todaro (1969: 140) does not necessarily confine himself to the context of developing countries. He cites, as an important precursor, a study of the United States by Schultz (1945:89 – 102) that discovered the paradox of parallel movements in farm prices and net migration from farms in the inter-war years. So much so that in

1932—the year with the lowest farm prices in thirty years—there occurred a net migration to farms! Accounting for relative prices did not change the picture either. The rise and fall in farm prices would generally be larger than in the other prices. Schultz concluded: 'Not prices, therefore, but the existence of job opportunities—the opportunity to migrate—takes farm people off farms or requires them to stay put' (1945:101).

It should be evident by now that the 'stigma' of being rooted in advanced industrial societies is not exclusive to the income or wage differential approach; Todaro's preferred job opportunity approach goes back to the same roots. His reliance on institutional and economic contexts different from those of developing countries was no less than his predecessors. The point to note is that wherever conditions of substantial unemployment prevail, employment becomes a determining influence on migration. As this is the conclusion emerging from Todaro's own account, the total disregard by him and others (see the often-quoted survey by Greenwood 1975) of the work by Robinson does not seem to be justified.

As pointed out above, Todaro did not care too much about the institutional and economic context of a developing economy in identifying the prehistory of the Harris-Todaro model. The fact that Robinson did not present her analysis in the context of the underdeveloped economies of those days cannot be the reason for ignoring her contribution. Nor could she be accused of assuming full employment and a price-adjusting equilibrium, assumptions that Todaro decried in his predecessors.[2] Could it be that a specific interest in rural-urban migration led Todaro to disregard the Robinson model and to choose Schultz as the most important precursor? The answer is no, as is confirmed by Greenwood in his authoritative survey: 'While the Todaro model is applied specifically to rural-urban migration in less-developed countries, the model is general enough to apply to interregional migration in any country' (1975:403). Interestingly enough, Suits (1985) has actually applied the Harris-Todaro model to the case of the United States.

In short, despite stated intentions to the contrary, the Todaro hypothesis and the Harris-Todaro model are by no means specific to the context of developing countries. The disregard of the work by Robinson cannot, therefore, be attributed to this factor. It has more to do with the less-than-careful regard the migration model builders seem to have had for their antecedents. This is amply borne out by the fact that in building a straw man out of Lewis (1954), no attention was paid to Lewis (1958) which contained 'a similar formulation long before the Harris-Todaro article appeared'.[3]

JOAN ROBINSON'S STORY IN PERSPECTIVE

The essentials of Robinson's story are set out in her essay on 'Mobility of labour'. Basically the concern here is with her treatment of geographic mobility (Robinson 1937a:53 – 7).

Determinants of migration

Consider an origin i and a destination j, and let migration be denoted by M. In the Hicks sense, the incentive to migrate from i to j, denoted by $M_{i,j}$, is the relative real wage differential ($W_j/W_i<1$). What Robinson emphasises is the point that even if $W_j/W_i<1$, $M_{i,j}$ may still follow. The relative real wage is not the only incentive to migrate. The key incentive is a mix of employment and what happens to money wages. Suppose investment is increased in j, it will also increase employment, but not without affecting prices. As more labour is applied to the same equipment at the same wage, output per person falls and the cost per unit of output goes up.

Prices must rise to maintain profits (Robinson 1937b:56). Thus the original increase in employment following a rise in investment in j leads to a higher cost of living. To look at migration into *j* from *i,* Robinson distinguished two cases depending on what happens to the money wage. These will be considered in turn.

Constant money wage

In the case in which money wage stays unaltered in *j*, the effect of a higher cost of living is that the real wage falls. More important for the analysis here, it is lower relative to i, ($W_j/W_i<1$). The Hicksian prediction will then be that labour should move to i. But the opposite is closer to reality in terms of Robinson's analysis, as unemployment in *j* (U_j) is lower relative to unemployment in *i* (U_i). The incentive for migration undergoes a fundamental change (see Robinson 1937a: 54).

It is abundantly clear that Robinson had discarded the real wage differential hypothesis soon after it had been presented. The principal determinant of migration was the chance to be employed, not the real wage differential. The latter only provided 'a slight pull' or 'increase[d] the incentive to immigration' (Robinson 1937a:55). These theoretical assertions preceded the empirical observations of Schultz (1945) in the quest for alternatives to the price-adjusting migration hypothesis.[4]

Much has been made in the literature of the Kenyan experiment with a tripartite agreement whereby the employers consented to expand employment to ease the rising urban unemployment and the employees agreed to a wage standstill. Additional jobs were created, but the migration induced from rural areas as a result was so great that urban unemployment increasedrather than

decreased (Harbison 1967:183; ILO 1972:542; see also Lewis 1958:4). Todaro (1969:140 – 1) cited the experiment as a 'contemporary' illustration of his hypothesis, and it is common to perform comparative statics on the Harris-Todaro equilibrium to explain the Kenyan experiment (see Basu 1984:70; Fields 1987). However, it has not been suggested that the constant money wage case in the Robinson model also brings to the fore the significance of the chance to be employed. Immigration 'is likely to be on a relatively small scale so long as' this chance is small, and 'to set in with a rush if' this chance is very high (Robinson 1937a:54). The latter implies, however, that employment in *j* must be very high and the money wage cannot then remain the same in the Robinson world.

Rising money wage

An important consequence of falling unemployment in *j* is that the 'scales tip in favour of the workers', who act in the hope of protecting their real wages from erosion by demanding higher money wages. But the 'prices move ahead of wages as the horizon moves ahead of the traveller' (Robinson 1937b:59 – 61). Thus the higher money wage in *j* allows the purchase of fewer *j*-goods than before. In the meantime, migration from *i* to *j* continues in response to rising job opportunities. The incentive to migrate increases if the terms of trade of *j* improve *vis-à-vis i*. In sum the principal determining influence on migration is the chance to be employed, not the real wage differential. But a rise in money wage, leading to improved terms of trade, enhances the incentive to move by causing a real wage differential.

The hypotheses: Robinson versus Todaro

Very simply, the Robinson hypothesis states that migration M_{ij} is positively related to the chance to be employed in *j*. As was noted above, a higher chance to be employed in *j* is associated with higher migration from *i* in the Todaro hypothesis as well. Both dismiss the real wage differential approach. This is, however, all that there is in the nature of similarities. Important differences arise in the conceptualisation of the chance to be employed and the respective Keynesian and choice-theoretic assumptions.

The Lewis migration takes place just like the Hicks migration when

$$W_j/W_i > 1 \qquad (1)$$

where W_i is the subsistence wage in the traditional sector and W_j the wage in modern sector. The members of surplus labour in *i* who migrate to *j* have a 100 per cent chance of securing a job. This was the major problem that Todaro (1969, 1976) found with the Lewis (1954) model. High rates of urban unemployment observed in most underdeveloped countries in the1960s contradicted the full employment assumption in *j*. In the Harris-Todaro (1970) model, the consistency

with reality is achieved by assuming the chance to be employed in *j* at less than one. There now exists a quantity of unemployment, U_j. The rate of employment in *j*, $(1-U_j)$, is considered a reasonable proxy for the probability of finding a job. However, the assumption of unity probability is shifted from the destination *j* to the origin *i*. The assumption of an urban labour surplus was adopted, and that of rural surplus discarded, for the empirical reasons that became standard (see Kao *et al* 1964; Turnham 1971). W_j is exogenous, enforced by government fiat.

In contrast to Harris-Todaro employment probability $(1-U_j)$, the proxy indicated in the Robinson model for the chance to be employed is relative unemployment, U_j/U_i (Robinson 1937a:54). A distinctive feature of the Robinson model becomes immediately obvious: she does not assume full employment either in *j* or in *i*. Her model is thus closer to the double reality of urban unemployment and rural differentiation. There is incentive for M_{ij} when

$$U_j/U_i < 1 \tag{2}$$

The concept of relative unemployment is symmetric, incorporating push as well as pull effects. It is a proxy for the basic determinant of migration in Robinson's story—the chance of finding a job. The chance exists independently of whether or not W_j/W_i is greater than one. As was seen above, a constant money wage implying a fall in W_j/W_i did not deter the inflow of migrants. And a rise in the money wage entailing a rise in W_j/W_i increased the incentive to move. But the chance to be employed remains the basic determinant of migration.

In the Harris-Todaro model, the employment probability $(1-U_j)$ is multiplied by W_j to obtain the expected wage in *j*, *e* W_j. Given an employment probability of unity, the expected wage in *i* is W_i. Thus according to the Todaro hypothesis, migration takes place when

$$eW_j > W_i \tag{3a}$$

or

$$W_j(1-U_j) > W_i \tag{3b}$$

As W_j is assumed to be given parametrically above the market-clearing level, it is the function of U_j,a quantity, to equilibrate the migration flows. As in the models of Hicks and Lewis, migration continues to be a pull phenomenon, with the crucial difference that the pulling force is an expected rather than a real wage differential. It is easy to see that the representation by Robinson of the chance to be employed by relative unemploymentenables her to avoid conceiving of the migrant as an individual acting autonomously in pursuit of a specified maximand. In comparison, the Harris-Todaro model is essentially choice-theoretic. With the probability of employment assumed in *i* being 100 per cent, the individual is 'free to choose' the location of their employment.

On the basis of (3b) above, the Harris-Todaro equilibrium is given by

$$W_i = W_j(1-U_j) \tag{4}$$

so that

$$W_i/W_j = (1-U_j) \tag{5}$$

Equation (5) entails a perverse implication: the unemployment rates it predicts are a lot higher than those observed in the underdeveloped economies.[5] For example, the original 30 per cent wage gap assumed by Lewis (1954) would require U_j=23%. Later suggestion by Lewis (1967:42) of a wage gap of 100 per cent in the case of Nigeria would have required for equilibrium U_j=50%.[6] More recent estimates put the wage gap at even higher levels, requiring even higher levels of equilibrium unemployment. Todaro (1976:38) protests against this 'literal interpretation' and chooses to call it an illustrative device for the more important inverse relationship between expected wage differentials and the destination rates of unemployment.[7] However, the essential policy message of the Harris-Todaro model is to allow a market-clearing urban wage. Absorbing a migrant in a >job means an inflow of $1/(1-U_j)$ migrants, foregoing $W_i/(1-U_i)$ which is equal to W_j. The opportunity cost of labour is the market wage. As Sen (1984: 260) put it: 'The Invisible Hand strikes again!'

Robinson takes the view that 'even a small amount of unemployment is a severe obstacle to mobility'. Unemployment at destination, no matter how small, tends to build up resistance factors at the destination and reluctance factors for intending migrants. The resistance factors find recognition in the literature on international migration,[8] but discussions of them in the context of internal migration have not been numerous. Lipton (1964) is an example, as he talks of migrants coming up against growing resistance from those already in jobs, especially the unionised competitors. The effect of the presence of resistance factors is that the bourse hiring pattern assumed in the Harris-Todaro employment probability, with all those showing up equally likely to be selected, moves further away from realism.

Besides the resistance factors, unemployment—even at low rates—breeds uncertainty for the prospective migrant, conditioning in particular the movement of married persons. As Mazumdar (1983) has shown, supply price of family migrants is higher than that of individual migrants. In the Robinson scheme of things, an individual improves the chance of finding a job bymoving from a high unemployment origin to a low unemployment destination. But the materialisation of this chance is subject to the ability of the migrant to overcome uncertainty. Uncertainty and risk were ignored in the original Harris-Todaro model, with later extensions following suit. The migrant is envisaged to be neutral or indifferent between certain access to low wage work at origin and the unsure prospect at destination of either getting a high wage job or waiting for it. If the waiting involves open unemployment, as Todaro has it, the required equilibrium unemployment is blown up to worse than the worst of all possible worlds. The

waiting could also be in what Fields called the 'murky sector' (Fields 1975:171 – 6), lowering as a result the predicted equilibrium unemployment. It has also the effect of reducing the uncertainty associated with unemployment. However, Cole and Sanders (1986:572, n8) find Fields' assumption that unemployment equilibrates murky as well as non-murky labour markets to be invalid. There is, according to them (Cole and Sanders 1985:488), a distinct stream of migrants heading only for the 'urban subsistence sector', where the probability of employment is unity.[9]

What has been termed as the 'new economics of migration' entirely shuns the Harris-Todaro risk-neutrality and looks at migration as a manifestation of risk aversion (Stark and Levhari 1982:191). It views it as a 'calculated strategy' rather than 'an act of desperation or boundless optimism' (Stark and Bloom 1985:175). Shifting the unit of analysis from the individual to the family does not make any difference. For the family, migration reflects 'risk-reducing portfolio diversification of income sources' (Stark 1982:67). Indeed, '*the familial* decision-making process leading to migration' is singled out as a major plank in the future research programme to deal with 'the weakness of migration theory' (Stark 1985: 1).

A few comparisons between these recent approaches and that of Joan Robinson are in order. First, following Keynes, Robinson's emphasis is on uncertainty rather than calculable risk (Robinson 1973a:4). The suggestion in her approach is that unemployment and uncertainty move in parallel fashion, but the latter declines more slowly than the former. Hence the observation that even low unemployment cohabits with uncertainty when resistance factors are present at the destination. This inhibits married migrants from being the plungers. In the 'new economics of migration' by contrast, the initial risk of unemployment in urban destinations is very high, but it declines over time and in any case is lower than the risks involved in agricultural production at rural origins. Further, if trading between immigrants at destinations is substantial, the earlier migrants might see gain in helping later migrants. Such networking would be a possible counter to Robinson's suggestion that the 'natives will look askance at immigrants' (Robinson 1937a:54). Further still, married persons taking the migration decisions as family members may not be as constrained as thought by Robinson.

On the whole, the two perspectives differ at the level of analysis. While Robinson's is a macro view, the 'new economics of migration' is concerned with micro decision-making. It raises the level of analysis from the individual migrant of the Harris-Todaro world to the migrant in a family, but is rooted more deeply in the choice-theoretic tradition. Second, the Cole-Sanders model incorporates the push factors in the determination of subsistence-subsistence migration, i.e. migration from the rural subsistence sector to the urban subsistence sector. The recognition of push factors brings their model closer to Robinson's approach, but the assumption of unit probability of employment in both subsistence sectors and

the consequent equilibrating of wages vitiates against the quantity-adjusting basis of her approach as well as that of Todaro. Actually, Todaro is strongly critical of the dual stream of migrants, finding it artificial (Todaro 1986:566). Third, the burden of the Todaro tradition was to show that migration into the cities of the underdeveloped countries was privately rational but socially costly. The message of the Robinson approach seems to suggest the social desirability of migration, especially in the context of making the economy more flexible (Robinson 1943:491). Cole and Sanders, as well as the 'new economics of migration' do not find migration to be undesirable. The latter, however, is more concerned to show how private rationality coincides with social desirability. To Robinson, migration seems to be an important part of the process of structural change. This entails a discussion of the effects of migration, which is the topic of the next section.

Consequences of migration

There has existed for a long time a consensus on the relative neglect of research, theoretical and empirical, on the consequences of internal migration. As late as 1985, Stark and Bloom noted 'a surprising lack of empirical work on the effects of labour migration' on destination as well as origin locations (Stark and Bloom 1985:177). As early as 1962, Sjaastad—pioneer of work on migration as investment in human capital and in whose tradition Todaro placed himself—noticed for the first time the predominance in migration research of causes and the neglect of consequences. In his influential survey in the mid-1970s, Greenwood (1975:397) re-confirmed Sjaastad's conclusion. A detailed survey by Todaro (1976:77) also identified 'a major and persistent knowledge gap' regarding the consequences of migration.

In the light of these comments, it is significant that a quarter century before Sjaastad published his seminal paper, Robinson presented a number of hypotheses on the consequences of migration. She wrote:

> Immigration will lead to a further increase in the level of employment [at destination]. Unemployed workers must make some expenditure, and part of their expenditure will fall on local goodswherever they may be at the moment. Immigration will therefore lead to a further increase of effective demand [at destination], and a reduction of effective demand [at origin]. Even if it is the rule that the unemployed are supported by a levy on the inhabitants of the centre where they happen to be, at any moment an influx of workers into a centre is likely to increase expenditure there, because it will cause a transfer to them of income from wealthier inhabitants of the centre. In other centres both employment and unemployment will decline as the result of emigration. Moreover, if the alteration in the geographical distribution of population is expected to be permanent, there will be a

> secondary inducement to increased investment [at destination], particularly in house building, accompanied by disinvestment in other centres. Thus any initial difference between the levels of activity in different centres will tend to be cumulative.
>
> (Robinson 1937a:55 – 6)

These remarks suggest what came to be recognised later as the two most important effects of migration: one related to the existence of simultaneity, the other to the cumulative nature of the consequences of migration.

Simultaneity

Not only is migration affected by employment, but migration itself affects employment. This is the unequivocal intent of the opening part of the above citation from Robinson. The cause is the rise in demand following immigration, as the immigrants must spend on something even when unemployed. Whether migration is self-financed,[10] supported by friends and relatives, or by a tax on wealthier residents at destination does not seem to matter as long as spending is located at the destination. Indeed, the tax alternative mentioned by her would have redistributive effects,[11] though the presence of a dole system is not very likely in the underdeveloped economies. Robinson also suggests symmetry of effects at destination and origin in regard to employment: an increase at destination corresponds to a decrease at origin. While unemployment also declines at origin as the unemployed leave, employment declines as the spending by emigrants stops. It is clear from the context that all emigrants do not immediately enter jobs at destination. Full employment is not assumed at destination; a waiting period is involved. However, employment increases as a direct result of the increased investment as well as due to the additional demand arising from migrant spending.

Some problems may be noted here. First, migrant spending is unlikely to be significant during the spell of unemployment in developing countries.Second, it is not only the unemployed that tend to migrate. Third, the existence of simultaneity creates problems for empirical work.

Cumulative effects

It was Myrdal who first articulated the case against the equilibrating effects of migration in his well known thesis of ‘circular and cumulative causation’. Migration, in terms of his analysis, is one of ‘the media through which the cumulative process evolves’, with the destination moving upwards and the origin downwards. The destination ‘where economic activity is expanding will attract net immigration’. Given the invariably age-selective nature of migration, ‘this

movement by itself tends to favour' the destination and 'disfavour' the origin (Myrdal 1957:27 – 9).

As is evident from the later part of Robinson's quotation above, she had predicted similar consequences for migration. She did not suggest that it is the most productive age groups who would migrate, but she did suggest that migrations might be expected to be permanent, and that increased investment was destined to attract migrants who will provide 'a secondary inducement to increased investment' so that 'any initial difference' between the destination and the origin 'will tend to become cumulative'.

MIGRATION AND JOAN ROBINSON'S LATER WORK

The Todaro hypothesis and the Harris-Todaro model contain a simple but powerful insight, namely that rural-to-urban migration continues despite rising urban unemployment. The migration decision is privately rational but socially costly. With the social desirability of migration in doubt, policy must address its costs—setting of urban wage above the market-clearing level. Urban employment policy based on a disequilibrium wage rate is a recipe for more urban unemployment. The object of this section is to ascertain Robinson's thinking on these important issues, and to attempt a reconstruction of the early migration story in the light of her later writings, especially those on development.

In her later writings, Robinson did not pay any special attention to migration. China stands out among the most important examples in modern times of positively discouraging rural-to-urban migration (see World Bank 1984: 98). In spite of her great interest in China (see Robinson 1977), she did not publish anything noteworthy on this aspect of the Chinese development. The nearest she came to dealing with the subject was to express a blind faith in planning:[12] 'planning should confer certain advantages in coping with the economic and social problems arising from the large-scale migration of labour' (Robinson and Adler 1958:15). In the case of Sri Lanka, she againemphasised it as a planning problem, but was more explicit about the migration induced by industrial development.

Later, however, the emphasis on population growth grows stronger. There are pull as well as push factors at work.

> A sharp rise in the rate of growth of numbers has set in in most of the countries of the Third World, and, in almost all of them, employment has been growing over the last twenty years less rapidly than population. Agriculture fails to provide even the barest livelihood for new generations of would-be cultivators, while the number of jobs in regular industry and commerce expands slowly. A flow of dispossessed families has drifted into shantytowns and slums or on to the streets of cities.
>
> (Robinson 1979a:5 – 6)

This is what has been described above as the murky or urban traditional sector, and is more commonly known as the informal sector. She does not see migrants into this sector as disguised unemployed, but

> as self-employed, with minute quantities of capital, providing communal and personal services to each other and to their prosperous neighbours, or working in small family businesses or for any amongst them who have acquired enough finance to employ those more wretched than themselves.
>
> (Robinson 1979a:6)

As these workers 'are obliged to provide services at a cheap rate', other urban residents find that it 'increases the purchasing power of money' for them. The result is not marginalisation but

> a conflict of interest within the formal sphere. Managers of department stores and supermarkets object to petty traders who take custom from them while manufacturers find that they provide an outlet for their products at very low profit margins and tap a level of the market that posh and expensive shops cannot reach.
>
> (Robinson 1979a:7)[13]

Thus Robinson did not believe that urban surplus labour emerged as a sudden result of some policies. Migration, with all its problems, was part of the process of development.[14] According to her, a 'civilisation of any level of complexity cannot develop unless it is possible to get a surplus from agriculture, which will feed city life and feed industry'. Migration, it may be inferred, has been an inevitable concomitant of the 'method of getting a surplus from agriculture which enabled civilisation to develop. If there hadbeen social justice and equality there would have been no civilisation' (Robinson 1970:37).

The question is not the desirability or otherwise of migration. In Robinson's framework, if there is a chance for the workers to be employed, they will move to the towns. Indeed, two parallel developments make Robinson's framework more relevant to the developing countries of today. On the one hand, high urban unemployment rates have become common. On the other hand, the rural sector has seen the emergence of wage labour, which also suggests the possibility of open rural unemployment (see P. Bardhan 1979:486 – 7, for evidence from India). Robinson's own tone was pessimistic:

> The general effect of the remnants of quasi-feudalism that remain in the Third World today is to retard both the growth of output and the development of a humane society. Total production and the transfer of product from agriculture to the rest of the economy are stagnant or grow very slowly. The spiral interaction of agriculture feeding the industry and

> industry equipping agriculture fails to take off. Employment, which distributes claims upon a share of the product within agriculture, grows more slowly than the population. Redundant workers are expelled from the rural economy before the industry has grown up to absorb them.
>
> (Robinson 1979a:47)

These developments indicate that the Robinson hypothesis of relative unemployment may be reflecting the reality of migration better than the Harris-Todaro model, as the latter does not allow unemployment at origin. The policy implication is to recognise the mutually supportive nature of agriculture and industry,[15] as opposed to the Harris-Todaro prescription for a large reduction in urban wages. Although industrial development was considered by her as the major objective of development, she told the Sri Lankans that 'the amount of direct employment that can be offered by organised industry is not great', limited as it is by 'the potential investment, and particularly the foreign exchange available for investment', not to speak of the 'difficulties of drawing up schemes'.[16]

Furthermore, 'it can never be right to pick up schemes in terms of the employment that they offer' (Robinson 1959:49). This is what the Kenyan experiment referred to above was all about—to make work. There are important reasons, according to Robinson, for not doing so. Expounding them before a Chinese audience, she pointed out that the argument was that employing more labour would require an allocation out of the additional output flow for the consumption of the newly employed. After the latter move to the cities, it is difficult to prevent the consumption of those remaining in the rural sector from rising and to force them to keep providingfood for those who migrate. In addition, the migrant in the city who works in a factory tends to consume more than the average rural resident. This is also part of the incentives-mix for moving to the city.[17]

Underlying the above argument is the concept of 'the inflation barrier'. In fact, her migration analysis of the 1930s rested on the price implications of increasing investment and employment. The latter is an 'area where Robinson's point of view did not change throughout her writings' (Asimakopulos 1984:389). Therefore, to reconstruct Robinson's migration story, it is important to spell out the concept of the inflation barrier. In what Robinson called a 'miserable economy' the inflation barrier sets limits on accumulation through the need to keep the workers 'fed if work is to be done' (Robinson 1956:356). In this economy,

> labour is unorganised and there is a mass of permanent quasi-unemployment among landless peasants or younger sons of small traders who are living on their families. The inflation barrier is reached when the level of real wages is so low as to impair the efficiency of workers, so that the employers themselves offer a rise in money wages to counteract a rise

> in prices. Peasants can support life (more or less) when living in their villages at a lower level of consumption than is required for the bare physiological minimum of an industrial worker.
>
> (Robinson 1956:49)

The concept is akin to the corn wage of the Classical model—'indispensable for the analysis of planned industrialisation of a backward over-populated country' (Robinson 1962:80)—or the Lewis-type constant real wage of the modern sector, which is nothing but rural subsistence topped up for the minimal requirements in the urban environment. There is, however, an important difference; Robinson does not ignore the terms of trade. She pointed out that the money wage deflated by a cost of living index gives a real wage different from that arrived at by a general output deflator. What constitutes the real wage in the eyes of the worker is not the cost of his labour to the employer (Robinson 1956:356).[18]

Productivity in manufactures and agriculture is different; the interests of town and country conflict; both supply and demand for food are price-inelastic. 'An increase in industrial employment, which raises demand in the cities, is liable to bring about a sharp rise in the price of foodstuffs. On the other tack, a good crop when demand is constant, actually reduces the purchasing power of rural income' (Robinson 1979:14 – 15).

With these insights from Robinson's later writings, it is possible to reconstruct the migration story. Let i be the rural/agricultural sector and j the urban/industrial sector in an underdeveloped economy. Industrialisation in j has just begun and investment goes up to support it. The increasedemployment opportunities in j will attract migration of workers from i. The money wages which it is necessary for employers to offer are held down by an elastic supply of labour, accustomed to very low earnings, from the overpopulated countryside (Robinson 1946 – 7:105). Robinson's basic argument is that the response of money wage rates to increased employment is sluggish. It lags behind the rise in industrial productivity. The case of the constant money wage occurs in an underdeveloped context if the i sector is overpopulated. As the cost of living rises with increased employment, real wage falls. In spite of this, migration into j continues. In an underdeveloped economy embarking on industrialisation, the cost of living rises not because of increased unit cost but because demand for food goes up in j following a rise in employment. Food is an exportable of i, with an inelastic supply. The demand for food is also price inelastic. Prices rise because of low supply and demand elasticities of food, not due to the need to maintain profits.

Increased employment in j leads to a rise in the cost of living as well as deterioration in terms of trade. This means that average real income in i relative to j has risen, and so has the level with reference to which the money wage was determined in j in the first place. Money wages in j have to rise. The case is described by Robinson as

> a low-level bastard golden age where real wages are at the subsistence minimum, so that any rise in the price of food-stuffs (due, say, to a harvest failure), or any attempt to increase the rate of accumulation, forces employers to offer higher money-wage rates in order to enable their workers to live.
>
> (Robinson 1962:70)

It may be recalled that a rise in the money wage in j in the developed-country context led to a rise in the real wage relative to i as imports from the latter became cheaper. In her words, 'the favourable effect upon the terms of trade of raising money wages may outbalance the tendency for increased home employment to lower real wages' (Robinson 1937a:55, n1). In an underdeveloped economy, increased accumulation and employment in j lead to higher demand for food imported from i. Higher money wages are thrust on employers 'to enable the workers to live'. She distinguished the setting up of an inflation barrier between the developed and the developing countries as follows:

> When labour is organised, irresistible demands for higher money-wage rates set an 'inflation barrier' to an unacceptable share of profits while, even in South Korea, real wages cannot be kept permanently below the level which permits workers to preserve their capacity to toil and bring up their families.
>
> (Robinson 1979b:xxiii)

What happens to the money wage is crucial to Robinson's analysis.[19]

Furthermore, the terms of trade are not allowed to move in the favour of i.[20] 'In the Third World, where industrialisation is taking place mainly in an agricultural setting, prices are manipulated politically' (Robinson 1979a:15). What it ensures in effect is a constant real wage in j. Workers will continue to pour into j from i so long as accumulation and employment continue to rise in j, despite a constant real wage, i.e. wage in terms of food. Accumulation is affected when the product wage rises and eats into profits. This is counteracted by the function that unemployment performs in the underdeveloped economies, which 'is not to keep money wages from rising but to keep real wages down to a starvation level' (Robinson 1974:221).

CONCLUSION

This essay has outlined the analysis Joan Robinson carried out in 1937 on migration, and placed it in the context of the literature on migration and development. It was pointed out that Todaro and the 'Todaro-generated conventional wisdom' (Rogers and Williamson 1982:473) took no account of her contribution. It was shown that Robinson was concerned, like Todaro, to provide

an alternative to the wage differential theory of migration. Her analysis, therefore, is an important precursor of the Harris-Todaro tradition. The essay also discussed Robinson's insight that not only does employment influence migration, the reverse is also true. Similarly, she thought that the effects of migration would be cumulative. While the effects of migration have been noted in the development and migration literature, they have not been studied as intensively and extensively as the determinants of migration.

Todaro recognised Schultz among his important precursors. This is consistent with his rejection of disguised unemployment, his adherence to the human capital approach, and the view of migration as the rational decision of an expected-income maximising individual. The privately rational decision, however, imposes social costs. The implication is to get the prices right and to re-fix the policy focus through downward wage flexibility. This is the most important conclusion of the Harris-Todaro model.

The Robinson hypothesis as seen in the light of her later emphasis on capital accumulation appears to be running into a cul-de-sac. If unemployment is rising rather than falling at urban destinations, why should migration continue, as it apparently has, in the developing world? In Robinson's framework, if investment is increased at destination, employment should rise and, relative to the origin, unemployment must fall for migration to continue. The Harris-Todaro explanation is neat: migration continues despite rising unemployment because the expected wage differential continues to be positive, kept up by an exogenously maintained urban wage. It may be correct to criticise the model for assuming unreal individualswho have behind them 'a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed' (Keynes, quoted in Robinson 1979d:xv-xvi). But it does not resolve the dilemma. Similar criticism in the United States of 'the vision of the Southern worker thumbing through his present value tables' was said to be beside the point (Bowles 1970:357).[21]

The proper course seems to be to follow the logic of Robinson's early work and concentrate research effort, in the development context, on so much that is unknown about the rural origin, which the Harris-Todaro model takes for granted. The general impression is that urban unemployment rates are higher than rural unemployment rates (World Bank 1984:97). Contrary to the Robinson hypothesis, migration is taking place from low-unemployment origins to high-unemployment destinations. However, the emergence of wage labour in the rural areas leads one to suspect that the unemployment rates there may not be that low. P.Bardhan (1979) provides some indication of this effect. Furthermore, the Harris-Todaro model fudges the question of how urban wage is determined. In Robinson's work, the concept of inflation barrier suggests an endogenous mechanism for keeping the urban wage above a strangulating equilibrium. The non-existence of such a mechanism at the rural origin may be reinforcing the incentive to move.

Joan Robinson's story starts with an increase in autonomous investment. As she never endorsed investment for jobs *per se,* of which the Kenyan experiment quoted by Todaro was an apt example, the concept of investment would have to be more carefully formulated. For developing countries, she made a distinction between 'the demand for labour generated by a stock of capital after it has been created' and 'the effect on employment while investment is going on' (Robinson 1979:115). The latter is the typical Keynesian investment, which may attract, in the short run, unskilled and uneducated migrants. Over the medium-term, the former is more likely to attract workers with some skill or education. Investment causing differential access to social and physical infrastructure itself becomes an important cause of migration. More important, as 'it is not easy to get the cultivator to exchange his crops for money when there is nothing for him to spend money on' (Robinson 1979a:16), a policy of correct farm prices would be defeated by not-uncommon failures to invest in incentives goods (Sender and Smith 1986:123).

The message of Joan Robinson's analysis is that wherever investment is undertaken, giving rise to the prospects of more jobs, workers will move to them. But 'the object of investment is to create industrial capacity and raise productivity of labour, not to "make work" for the unemployed'. As migration is nowhere stated to be socially undesirable, the policy message is not the Harris-Todaro wage flexibility in towns, but 'maximum possible increase in employment per unit of investment' through 'a rational plan' (Robinson1979a: 116). But her work never specifies what a rational plan is, as the next essay on China[22] makes it abundantly clear.

NOTES

1 Part of author's doctoral dissertation at the University of Cambridge, supervised by Geoff Harcourt.

2 In a sense, as Robinson wrote in the context of the depressed regions in Britain, she was dealing with an experience that would constitute an invaluable backdrop for the 'Pioneers in Development' (Singer 1984:276 – 8).

3 Sukhamoy Chakravarty's letter to the author, 13 July 1987. Lewis (1958:44) stated what became the essence of the Harris-Todaro model in these words:

> It is very difficult to know how to cope with this increase in urban unemployment. The normal way to cope with unemployment is to provide work, but this is no solution in this case. On the contrary it merely aggravates the problem, because the more work you provide in the towns the more people will drift into the towns, and there is no certainty that you can win the race…. No one ought to say that he knows how to cure unemployment in this situation, or that he has a

> means to providing sufficient employment rapidly for all those who are unemployed in the towns.

4 The so-called Oxford Studies, published a year after Robinson (1937a) and well before Schultz (1945), found a high correlation between the rate of emigration at origin and the difference between the rates of employment between origin and destination. See Makower *et al.* (1938).

5 This was noted, among others, by Fields (1975).

6 A 30 per cent wage gap means that W_j=130, if W_i=100. Plugging these values in Equation (5) and solving for U_j gives 100/130=$(1-U_j)$ or U_j=1−(100/130) =23%.

7 On the issue of irrationally high equilibrium unemployment rates, see the encounter between Cole and Sanders (1985:483 – 4) and Todaro (1986:567).

8 As Robinson noted in the case of international migration, the destination residents treat immigrants 'as inferior and resent their presence'. Their 'protests against unemployment' tend 'to be deflected into protests against immigrants'. The unions fail to organise immigrants 'against the xenophobia of their native rank and file'. Most important, '[all] this, mutatis mutandis, is true of the home-born non-white population of the USA' (Robinson and Eatwell 1973:309). The reference to the non-white migration within the United States, comparisons to which are often drawn in the development literature, suggests that her elaboration of the resentment factors at international destinations applies to internal destinations as well.

9 As an afterthought, Todaro made a footnoted reference to risk and uncertainty in a later writing (1976:31, n1).

10 The need to finance migration makes it selective, with labour market imperfections becoming less significant than capital market imperfections in regard to the allocation of labour. See Lipton (1982).

11 Stark (1982) notes that most past studies predicted worst distributive impacts of migration.

12 'Planning should confer certain advantages in coping with the economic and social problems arising from the large-scale migration of labour' (Robinson and Adler 1958:15). In the case of Sri Lanka, she again emphasised it as a planning problem, but was more explicit about the migration induced by industrial development (Robinson 1959:45).

13 The possibility of urban employment other than that provided by the capitalist enterprise had been noted by Robinson in the early 1950s (see Robinson 1952: 110).

14 Kalecki regarded rural-urban migration as among the important causes of a rise in living standards in the process of development (1955).

15 See also the Joan Robinson Collection at King's College, Cambridge (JVRC, iii/9).

16 Sri Lanka is a small economy. In the large-economy case of India, she stated:

> The main problem is the growth of the potential labour force relatively to the stock of means of production. Even if all the bright hopes of the successive five-year plans had been fulfilled, the growth of industry could have made only a small dint in this problem. The main means of production is the cultivable land.
>
> (Robinson 1974:221)

17 See the Joan Robinson Collection at King's College, Cambridge (JVRC, iii/5).

18 For a similar distinction between the product wage rate and real wage based on cost of living in the context of the relevance of the Lewis model to the Indian economy, see Chakravarty (1977:119 – 21).

19 The question has assumed new significance in view of the following claim by the 'new economics of migration':

> High 'institutionally determined' wages in urban labor markets in LDCs are thus not necessarily externally imposed on the reluctant employers by government legislation and trade unions. Indeed, they may result from endogenously determined strategies designed to maximise profits in dynamic settings.
>
> (Stark and Bloom 1985:175)

20 On the other hand, even when the terms of trade move in favour of *i*, as in China, *j*-ward migration has to be prevented. As Robinson noted: 'Indeed, in spite of all efforts at regulation, it is not possible altogether to eliminate infiltration into towns' (1973b:28).

21 T.P.Schultz (1982:118, n7) also suggests that migration modelling does not require such a restrictive behavioural assumption.

22 In the case of rural-urban migration in China, sometimes she suggests mechanisation to release labour for industry (Robinson 1954a:28 – 9; Robinson and Adler 1958:17 – 18). At others it 'is a contribution to increasing the attractiveness of rural life to the younger generation and checking the drift into cities' (Robinson 1973b:10).

REFERENCES

Asimakopulos, A. (1984) 'Joan Robinson and economic theory', *Banca Nazionale del Lavoro Quarterly Review,* 37, 381 – 409.

Bardhan, P. (1979) 'Wages and unemployment in a poor agrarian economy: a theoretical and empirical analysis', *Journal of Political Economy,* 87,479 – 501.

Barnum, H.N. and Sabot, R.H. (1977) 'Education, employment probabilities and rural-urban migration in Tanzania', *Oxford Bulletin of Economics and Statistics,* 39,109 – 26.

Basu, K. (1984) *The Less Developed Economy,* Oxford: Basil Blackwell.

Bowles, S. (1970) 'Migration as investment', *Review of Economics and Statistics* 52, 356 – 62.

Chakravarty, S. (1977) 'Reflections on the growth process in the Indian economy', in C.D.Wadhva (ed.) *Some Problems of India's Economic Policy,* New Delhi: Tata McGraw-Hill.

Cole, W.E. and Sanders, R.D. (1985) 'Internal migration and urbanisation in the third world', *American Economic Review,* 75,481 – 94.

Fields, G.S. (1975) 'Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs', *Journal of Development Economics,* 2, 165 – 87.

——(1987) 'Public policy and the labour market in developing countries', in D. Newberry and N.Stern (eds) *The Theory of Taxation for Developing Countries,* New York: Oxford University Press.

Greenwood, M.J. (1975) 'Research on internal migration in the United States: a survey', *Journal of Economic Literature,* 13, 397 – 433.

Harbison, E. (1967) 'The generation of employment in newly developing countries', in J.Sheffield (ed.) *Education, Employment and Rural Development,* Nairobi: East Asia Publishing House.

Harcourt, G.C. (1995) 'Joan Robinson: 1903 – 1983', *Economic Journal,* 105, 1228 – 43.

Harris, J.R. and Todaro, M.P. (1970) 'Migration, unemployment and development: a two-sector model analysis' *American Economic Review,* 60, 126 – 42.

Hicks, J.R. (1932) *The Theory of Wages,* London: Macmillan.

International Labour Office (1972) *Employment, Income and Equality: A Strategy for Increasing Productive Employment in Kenya,* Geneva: ILO.

Kalecki, M. (1955) 'The problem of financing economic development', *Indian Economic Review,* 2, 1 – 22.

Kao, C.H.C., Ansdell, K.R., Eicher, C.K. (1964) 'Disguised unemployment in agriculture: a survey', in C.K.Eicher and K.W.Witt (eds) *Agriculture in Economic Development,* New York: McGraw-Hill.

Lewis, W.A. (1954) 'Economic development with unlimited supplies of labour', *Manchester School of Economic and Social Studies,* 22, 139 – 91.

——(1958) 'Employment policy in an underdeveloped area', *Social and Economic Studies,* 1,42 – 54.

——(1967) *Reflections on Nigeria's Economic Growth,* Paris: OECD.

Lipton, M. (1964) 'Population, land and diminishing returns to agricultural labour', *Oxford Bulletin of Economics and Statistics,* 26, 123 – 57.

——(1982) 'Migration from rural areas of poor countries: the impact on rural productivity and income distribution', in R.H.Sabot (ed.) *Migration and the Labour Market in Developing Countries,* Boulder CO: West view Press.

Mazumdar, D. (1983) 'The rural-urban wage gap, migration and the working of urban labour market: an interpretation based on a study of the workers of Bombay city', *Indian Economic Review,* 18, 169 – 98.

Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions,* London: Methuen.

Owen, A.D.K. (1937) 'The social consequences of industrial transference', *Sociological Review,* 29, 331 – 54.

Parr, J.B. (1966) 'Outmigration and the depressed area problem', *Land Economics,* 42, 149 – 59.

Robinson, J. (1936) 'Mr Keynes and socialism', *Left Review,* 2, 853.

——(1937a) *Essays in the Theory of Employment,* London: Macmillan.

——(1937b) *Introduction to the Theory of Employment,* London: Macmillan.

——(1943) 'Do we need world markets?', *Listener,* 30, 490 – 1, 501.

——(1946 – 47) The pure theory of international trade', *Review of Economic Studies,* 14, 98 – 112.

——(1952) *The Rate of Interest and Other Essays,* London: Macmillan.

——(1954) *Letters from a Visitor to China,* Cambridge: Students' Bookshops. Based on letters published during 1953 – 4 in *Monthly Review,* 5, 302 – 10, 397 – 407, 477 – 80, 536 – 43.

——(1956) *The Accumulation of Capital,* 1st edn, London: Macmillan.

——(1959) 'Economic possibilities of Ceylon', in *Papers by Visiting Economists,* Colombo: National Planning Council.

——(1962) *Essays in the Theory of Economic Growth,* London: Macmillan.

——(1970) 'Society and economics in China today', in R.Jungk *et al., China and the West: Mankind Evolving,* London: Granstone Press.
——(1973a) 'What has become of the Keynesian revolution?', in J.Robinson (ed.) *After Keynes,* Oxford: Blackwell.
——(1973b) *Economic Management in China, 1972,* London: Anglo-Chinese Educational Institute.
——(1974) 'Inflation west and east', in J.Robinson (1979) *Collected Economic Papers,* vol. 5, Oxford: Blackwell.
——(1977) *Reports from China, 1953 – 76,* London: Anglo-Chinese Educational Institute.
——(1979a) *Aspects of Development and Underdevelopment,* Cambridge: Cambridge University Press.
——(1979b) 'Introduction, 1978', in J.Robinson, *The Generalisation of General Theory and Other Essays,* 2nd edn of Robinson (1952), London: Macmillan. .
Robinson, J. and Adler, S. (1958) *China: An Economic Perspective,* London: Fabian International Bureau, Fabian Tract no. 314.
Robinson, J. and Eatwell, J. (1973) *An Introduction to Modern Economics,* London: McGraw-Hill.
Rogers, A. and Williamson, J.G. (1982), 'Migration, urbanisation and third world development', *Economic Development and Cultural Change,* 30:463 – 82.
Schultz, T.P. (1982) 'Notes on the estimation of migration decision functions', in R. H.Sabot (ed.) *Migration and the Labour Market in Developing Countries,* Boulder CO: Westview Press.
Schultz, T.W. (1945) *Agriculture in an Unstable Economy,* New York: McGraw-Hill.
Sen, A. (1984) *Resources, Values and Development,* Oxford: Blackwell.
Sender, J. and Smith, S. (1986) *The Development of Capitalism in Africa,* London: Methuen.
Singer, H. (1984) 'The terms of trade controversy and the evolution of soft financing', in G.M.Meier and D.Seers (eds) *Pioneers in Development,* New York: Oxford University Press.
Sjaastad, L.A. (1962) The costs and returns of human migration', *Journal of Political Economy,* 70, 80 – 93.
Stark, O. (1982) 'Research on rural-to-urban migration in LDCs', *World Development,* 10, 63 – 70.
——(1985) 'Introduction', *Journal of Development Economics,* 17, 1 – 3.
Stark, O. and Bloom, D.E. (1985) 'The new economics of migration', *American Economic Review,* Papers and Proceedings 75, 173 – 8.
Stark, O. and Levhari, D. (1982) 'On migration and risk in LDCs', *Economic Development and Cultural Change,* 31, 191 – 6.
Suits, D.B. (1985) 'US farm migration: an application of the Harris-Todaro model', *Economic Development and Cultural Change,* 33, 815 – 28.
Suzuki, K. (1970) 'Variation of regional population in Japan', *Journal of Regional Science,* 10,335 – 51.
Todaro, M.P. (1969 'A model of labor migration and urban unemployment in less developed countries' , *American Economic Review,* 59, 138 – 48.
——(1971) 'Income expectations, rural-urban migration and employment in Africa', *International Labour Review,* 104, 387 – 413.
——(1976) *Internal Migration in Developing Countries,* Geneva: ILO.

——(1986) 'Internal migration and urbanization in the third world: comment', *American Economic Review,* 76, 566 – 9.

Turnham, D. (1971) *The Employment Problem in Less Developed Countries,* Paris: OECD.

World Bank (1984) *World Development Report,* New York: Oxford University Press.

Part II

ECONOMIC THEORY AND APPLIED ANALYSIS

18
THE CAMBRIDGE THEORY OF DISTRIBUTION IN THE SHORT PERIOD
An open economy approach

Jorge Thompson Araujo[1]

INTRODUCTION

The Cambridge, or post-Keynesian, or neo-Keynesian theory of distribution is the result of the original attempts of Cambridge economists such as N. Kaldor, Joan Robinson, R.F.Kahn and L.L.Pasinetti to extend the Keynesian-Kaleckian principle of effective demand to the long period. Its most celebrated result is the so-called Cambridge theorem of the rate of profits. If the economy grows at Harrod's natural rate of growth (g_n), money wages are entirely consumed and aggregate profits are saved at the constant rate s_p, the long-period, uniform rate of profits (r) will be $r=g_n/s_p$, the Cambridge theorem in the Kaldorian version.

Famous as it is, the Cambridge *theorem* is a particular result of the Cambridge *theory,* so that the latter is not to be reduced to the former. The theory is based on the general idea that the principle of effective demand may govern not only the level of aggregate output, but also the distribution of income between wages and profits, or, in the Pasinettian version, between workers and capitalists. Whether effective demand rules the level and/or the distribution of income depends on the unemployment rate, the level of capacity utilisation, the existence of overhead labour and the flexibility of profit margins with respect to money wages (Araujo 1994, 1995). The main implication, for our purposes, of this relationship between effective demand and income distribution is that *neither absorption policies (policies that affect the level of demand) nor switching policies (policies that affect the composition of demand) are distributively neutral*.

The purpose of this paper is to present a short-period version of the Cambridge theory, within an open economy approach. The choice of the object is justified on the following grounds:

1 The early versions of the theory, both in Keynes' *The General Theory* and in Hahn's (1951) article (recently praised by Solow [1992]), were developed within a short-period context, with a given stock of capital goods. Despite this,

most of the voluminous literature on the subject (with a number of relevant exceptions[2]) is concentrated on long-period matters;

2 our open economy approach allows for the introduction of nominal and real exchange rates, that have important distributional impacts;

3 a short-period framework, from a Keynesian perspective, is the traditional set-up for analysing policy matters, which is one of the major purposes of this paper;

4 this paper links together in an integrated whole some frequent themes in heterodox macroeconomics, such as income distribution, money-supply endogeneity and real-wage resistance, generating interesting novel insights.

A short-period theory of distribution, however, has to be concerned with the *share* of profits in income rather than with the *rate* of profits, which loses most of its significance as a macroeconomic variable, except as an average index of capacity utilisation. Outside the realm of a long-period equilibrium, no clear meaning can be attached to the notion of 'the' rate of profits:

> The rate of profit on capital, in a short-period situation, is an even more foggy notion than the level of profits earned by given equipment, for to express profits as a rate we must know the value of capital.
>
> (Robinson 1956:190)

The value of capital has a precise meaning only in the context of a long-period equilibrium, in which productive capacity is already adjusted to the long-period level and composition of demand.

Bearing this warning in mind, we develop a model for the share of profits in an open economy framework with a floating exchange rate in the third section. Before that, we discuss the role of money-supply endogeneity in an open economy and how it affects the results of the standard Mundell-Fleming approach. In the fourth section, we deal with what we call the *Harcourt effect* of a *negative* multiplier effect of a higher investment demand within a fixed-exchange-rate version of the model of the previous section. The fifth section contains a brief analysis of the case of real-wage resistance in an open economy. The last section is devoted to some concluding remarks.

ENDOGENOUS MONEY SUPPLY IN A SMALL OPEN ECONOMY

There is a clear dichotomy in mainstream open economy macroeconomics regarding the money supply and the roles of fiscal and monetary policies under fixed and flexible exchange-rate regimes (Dornbusch 1980: ch. 10, ch. 11). Assuming perfect capital mobility, money supply in a *small* open economy under *fixed* exchange rates is entirely endogenous and the LM curve is perfectly elastic

at the level of the interest rate prevailing in the world capital markets. Monetary policy, interpreted as changes in the money stock decided by the monetary authorities, is completely ineffective under the circumstances described. Fiscal policy, in turn, is very effective, as increased government spending or reduced taxation do not raise the equilibrium—'world'—interest rate (which is taken as given by our small open economy), thus being entirely absorbed by a rise in income.

The opposite occurs under *flexible* exchange rates and perfect capital mobility. The conventional Mundell-Fleming model predicts that the monetary authorities regain control over the money supply as the exchange rate becomes the adjustment factor that 'corrects' any disequilibrium in the balance of payments—the money stock is, as in autarchy, *exogenously* determined. While monetary policy is now very effective, fiscal policy becomes completely ineffective: a fiscal expansion is compensated for by an exchange-rate appreciation that 'corrects' the balance of payments surplus that follows the initial rise in the interest rate—the higher government deficit is totally matched by a full crowding-out of net exports.

If the money stock is entirely demand-determined, the dichotomy between the two exchange-rate regimes disappears and both fiscal and monetary[3] policy may be effective (to varying degrees). We have now a *fixprice* model in the money market (Morishima 1992:174). As quantity-adjustment prevails in that market, there is no reason to suppose that the process of endogenous money creation will be affected by the prevailing exchange-rate regime (Arestis and Eichner 1988:10 – 15)[4]. What the exchange-rate regime affects is the ability of the monetary authorities to set the rate of interest (i.e. the 'price' variable) when the degree of international financial capital mobility is very high. Under those conditions, the monetary authorities will be more or less constrained in setting the 'official' rate of interest under fixed (flexible) exchange rates (Moore 1988:274).[5]

Mundell's (1961) article is particularly appropriate for a short-period, one-commodity analysis of this case, since monetary policy is assumed to take the form of shifts in *interest rates,* which is analytically equivalent to the assumption of money supply endogeneity.[6] As it will become clear shortly, the striking implication of this hypothesis is that a fiscal expansion will lead to an exchange-rate depreciation, and not an appreciation as assumed in models with money supply exogeneity (e.g. Dornbusch 1976). Therefore, if the Marshall-Lerner condition is met, there is no reason to expect a crowding-out of net exports to follow a rise in government spending.

INCOME DISTRIBUTION WITH FLEXIBLE EXCHANGE RATES AND FIXED MONEY WAGES IN THE SHORT PERIOD

The model to be discussed here descends directly from the contributions of Harcourt (1963, 1965, 1972), Asimakopulos (1969, 1970, 1975), and Harris (1974), among others. The opening of the economy to international trade and capital flows draws on previous works by Blecker (1989), Bhaduri and Marglin (1990), Sarantis (1990 – 1) and Taylor (1991). The main novelty of our approach is to integrate the analysis of exchange-rate determination by Mundell (1961) with the Cambridge distribution theory in the short period, assuming money supply endogeneity and the existence of overhead labour, at the same time that a simple framework for the discussion of policy matters is set up.

We follow Asimakopulos (1969, 1970) in supposing that the (given) stock of capital consists of accumulated output produced in the past. Accumulated past output, once incorporated in the capital stock, cannot be 'undone' and consumed. The structural form of the model for a small, open economy is presented below.

Technological requirements

(1a) $aY=M$, $a=M/Y$ (imported raw materials per unit of output); (1b) $bY=L_1$, $b=L_1/Y$ ('direct' labour per unit of output); (1c) $L_0+L_1=L^d$ (total demand for labour).

Product market

(2a) $u=Y/K$ (capacity utilisation); (2b) $I/K=I/K\ (i, u)$ (investment function); (2c) $pS=s\ (\omega)Y$ (private domestic saving function); (2d) $X=X\ (\theta, u^*)$ (exports function); (2e) $-p\gamma K+s\ (\omega)\ Y+ep^*aY-PX(\theta, u^*)=p(I/K)(i, u)\ K$ (product market equilibrium condition).

Value of the supply of output

$$(3a)\ pY = P + w\,(bY + L_0) + ep^*aY.$$

Pricing rule

$$(4a)\ p = (1 + m)(wb + ep^*a) = \psi(wb + ep^*a),\ \psi = 1 + m.$$

Real exchange rate

$$(5a)\ \theta = ep^*/p.$$

Capital inflows (in terms of domestic currency)

(6a) $F = F(i-i^*)$, $F' > 0$, $i \neq i^*$.

Exchange-rate determination

(7a) $pX(\theta, u^*) - ep^*aY = -F(i - i^*)$ (flexible exchange rates).

Endogenous money supply

(8a) $i = i_0$

Real-wage determination

(9a) $\omega = w/p$.

We have nine blocks of equations, with fifteen equations determining fifteen endogenous variables (M, L_1, L^d, Y, u, P, X, S, I, p, θ, F, e, i, ω) in terms of exogenous variables and parameters (L_0, K, γ, p^*, w, e_0, i_0, i^*, u^*; a, b, ψ, m), whose meaning is explained below.[7] In the first block, Equations (1a) and (1b) show the physical quantities of imported raw materials *(M)* and 'direct' labour (L_1), respectively, needed to produce a physical quantity Y of output. Constant returns to scale are assumed and coefficients a and b are given. Equation (1c) introduces overhead labour (L_0) as part of total labour demand *(L^d)*.

The second and third blocks display aggregate demand and supply conditions. Equation (2a) defines capacity utilisation *(u)* as the ratio between current output *(Y)* and accumulated past output *(K)*, supposing for simplicity a unitary full-capacity capital-output ratio.[8] Investment as a fraction of the capital stock is shown in Equation (2b) as an increasing function of u and a decreasing function of the domestic money interest rate i.[9] A standard Cambridge saving equation (2c) is assumed, where nominal saving *(pS)* depends on the real wage *(co)*. We postulate that higher real wages stimulate consumption, so that $s\hat{} < 0$ (Taylor 1991). Real exports X (equation (2.d)) depend positively on the world's average rate of capacity utilisation *(u*)* and on the economy's external price competitiveness *(θ)*. Equation (2e) shows the product market equilibrium condition in terms of domestic currency, so that imports in foreign currency $p^*a\,Y$ (with p^* as the rest-of-the-world's given average price level) must be multiplied by the exchange rate e (the domestic price of foreign currency). The real fiscal deficit γK is given and evaluated at the domestic price level p. Equation (3a) indicates that output has three distinct claimants: profits, wages and the cost of imported raw materials (we ignore, for simplicity, indirect taxes).

Pricing assumptions are shown in the fourth and fifth sets. The price level p is a proportion ψ of total average prime costs (Equation (4a)), where ψ is positively

related to the mark-up rate *m*. International price competitiveness *(θ)* is measured by the real exchange rate, as shown in Equation (5a).

The sixth and seventh sets introduce the overseas sector. Capital inflows are simply assumed to be an increasing function of the domestic *(i)* and foreign *(i*)* interest rate differential (Equation (6a)).[10] Imperfect capital mobility implies $i \neq i^*$, so that the monetary authorities have a degree of flexibility in setting the domestic rate. Equation (7a) displays a floating exchange-rate regime, indicating that the nominal exchange rate *e* floats in order to maintain balance-of-payments equilibrium[11]. It is taken for granted that the Marshall-Lerner condition holds, and J-curve effects are ignored. Following Meade and Vines (1988), we neglect interest payments in the balance of payments (which would not be acceptable in a longer-run perspective).

Equation (8a) represents the '*LM* side' of the economy, according to which the monetary authorities are unable to impose a given stock of money, but can and do fix the domestic interest rate. This hypothesis, now standard in several post-Keynesian models, is also present in the New Keynesian contributions of e.g. Meade and Vines (1988) and Weale *et al.* (1989).

Equation (9a) complements the pricing decisions block, presenting the real wage rate ω as the ratio of the fixed money-wage rate *w* and the price level *p*.

Our basic distributive variable is π, defined as the share of profits in the value of output *(P/p Y)*, which is a ratio of two endogenous variables, *P* and *pY*. Its governing equation results from (3a) and (4a) (Harris 1974):

$$\pi = [(\psi - 1)/\psi] - (\lambda_w L_0/\psi b K)(1/u) \tag{10}$$

so that π is a decreasing linear function of $(1/u)$, the inverse of the rate of capacity utilisation, and of λ_w, the share of wages in unit prime costs. The fact that $(1/u)$ tends to infinity as *u* approaches zero makes it necessary to define an interval of validity for (10) in order that the results are economically meaningful. With a unitary capital/output ratio (by construction), full-capacity utilisation provides an upper limit for π, whenever $u=1$:

$$(1/u)_{max} = (\psi - 1) b K/\lambda_w L_0 \tag{10a}$$

A natural lower limit for π is $\pi=0$, implying a maximum $(1/u)$:

$$(1/u)_{max} = (\psi - 1) b K/\lambda_w L_0 \tag{10b}$$

so that (10) can be represented diagrammatically as below in Figure 18.1.

Equation (10) shows that changes in income distribution may result from the 'demand side', i.e. $(1/u)$, or from the 'supply side' *(ψ, e, b)*, assuming L_0 and *K* to be invariant in the short period. 'Demand-side' changes are represented by *movements along* the π schedule, which may be called the *profits curve* (Rowthorn 1981). 'Supply-side' changes alter the *position and slope* of the profits curve. Point A on Figure 18.1 represents the profit share that would obtain

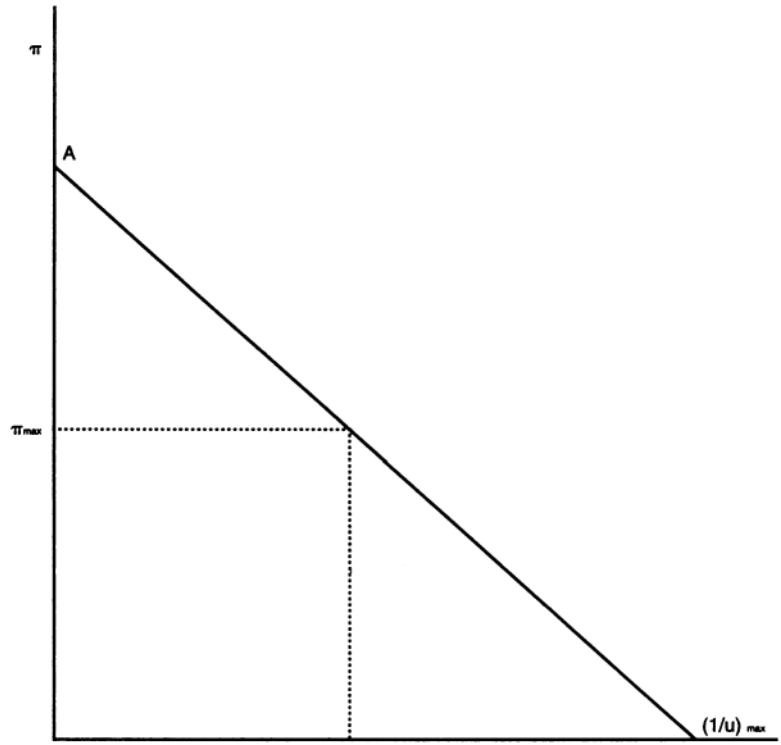

Figure 18.1 Diagrammatic representation of Equation (10)

if we neglected overhead labour: $\pi=(\psi-1)/\psi$, so that income distribution would be exclusively governed by the *degree of monopoly* existing in the economy (as captured by ψ). An exchange-rate devaluation, for instance, with a given degree of monopoly, would lead the profits curve to rotate anti-clockwise, with a fixed 'notional' vertical intercept at A.

It is convenient to re-express the product market equilibrium condition (2e) using an endogenous overall marginal propensity to save $\sigma(.)$:

$$\{s(\omega) + \theta a\}u - \gamma = \sigma(\omega, \theta)u - \gamma = g(i, u) + x(\theta, u^*) \qquad (11)$$

where $g=I/K$ and $x=X/K$. A higher real wage implies a decrease in $\sigma(.)$; higher competitiveness means higher real cost of imports, increasing external saving.[12] Those relationships imply that the partial derivatives σ^w and σ^θ are respectively negative and positive.

Money-supply endogeneity ensures that monetary and fiscal policies can affect the level of economic activity and, consequently, the profit share in an economy with a flexible exchange rate. As exchange rates fluctuate, income distribution effects also arise from the 'supply side', as the share of wages in unit

prime costs shifts in the opposite direction of the nominal exchange rate. Figure 18.2 illustrates the distributive impact of expansionary monetary and fiscal policies under floating rates, using a version of Mundell's (1961) diagram as reference. The IS curve results from the product market equilibrium condition (11), assuming that exchange-rate depreciations are expansionary. The BP curve is derived from the balance-of-payments equilibrium condition (7a), assuming that the Marshall-Lerner condition holds.

Both a fiscal and a monetary expansion (i.e. a larger budget deficit or a lower domestic interest rate) shift the IS to the right, moving the economy from point A to point B. At point B, there is a balance-of-payments deficit: at that level of output, the exchange rate is too low to ensure external balance. In the case of an expansionary fiscal policy, the balance-of-payments deficit is caused by an excess demand for imports—the capital account is not affected as, with endogenous money supply, fiscal policy *per se* does not change the interest rate. The ensuing exchange-rate depreciation shifts the economy to point C, further stimulating output and eliminating the external disequilibrium through an adjustment in the *current* account. A higher profit share is associated with a higher demand as a result of the depreciation, which causes the profits curve (0) to rotate downwards to its new position (1). The new share of profits is π_1.

A monetary expansion—an administrative decision to reduce the domestic interest rate from i_0 to i_1—has a slightly more complex effect. The adjustment process from point A to point C is as above, but does not stop there. A lower domestic interest rate affects the position of the BP curve—for a given output, the exchange rate must depreciate to compensate for the resulting capital outflow. Adjustment now involves the *capital* account as well—the exchange rate must depreciate *by more* than in the case of a fiscal expansion in order to counterbalance the rise in imports *and* the outflow of capital. Equilibrium will be reached at point D, with a higher u than thatwhich results from the fiscal expansion. The profits curve is subject to a further downward rotation, up to position (2)—the equilibrium share of profits is $\pi_2 > \pi_1$.

A rise in the foreign interest rate (i^*) to i_1^* which is *not* followed by a rise in the domestic rate also favours the level of capacity utilisation and increases the share of profits, both through 'demand-side' and 'supply-side' effects. A higher i^* shifts BP upwards: at the same level of output, a higher exchange rate is required in order to stimulate net exports sufficiently to offset the capital outflows that follow the increase in i^*. Higher net exports(ruling out contractionary exchange rate depreciation) produce a favourable 'demand-side' impact on the share of profits, which is reinforced by the reduction in the share of wages in unit prime costs caused by the depreciation. This analysis, however, is strongly based on the assumption of *no interest payments abroad,* notwithstanding its confirming a standard Mundell-Fleming outcome.

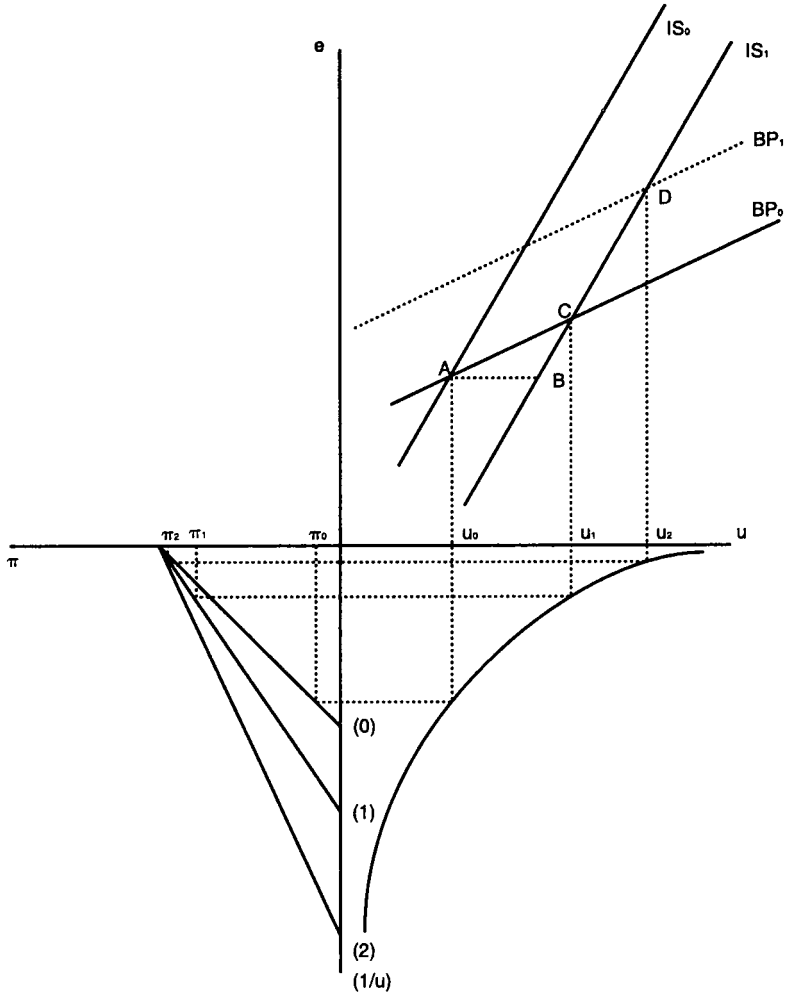

Figure 18.2 Distributive impact of expansionary monetary and fiscal policies under floating rates

A FIXED EXCHANGE RATE REGIME AND THE 'HARCOURT EFFECT' IN AN OPEN ECONOMY

A real exchange rate devaluation may be contractionary or expansionary in our model, because aggregate demand depends on the real-wage rate (Krugman and Taylor 1978). It is convenient to work with a fixed nominal exchange rate regime, so that Equation (7a) is no longer valid as e becomes a policy variable. From (11), we see that the impact of a higher real exchange rate on the level of capacity utilisation is measured by:

$$\frac{du}{d\theta} = \frac{x_\theta - \sigma_\theta u}{\sigma(\omega, \theta) - g_u}$$

Stability of saving-investment equilibrium requires saving to be more responsive to changes in u than investment, so that $\sigma(\omega, \theta) > g_u$ by assumption. *Contractionary* real devaluation ($du/d\theta < 0$) requires:

$$x_\theta < \sigma_\theta u \quad \therefore \quad \frac{dx}{d\theta}\frac{\theta}{x} < \sigma_\theta u$$

That is,

$$\eta \frac{x}{u} < \sigma_\theta \theta \tag{12}$$

where η is the real exchange rate elasticity of the export volume as a fraction of the capital stock.

Equation (12) replicates Taylor's (1991:145) condition for contractionary real devaluation. A nominal exchange-rate devaluation, from (4a), leads to some increase in the price level[13], reducing the real-wage rate and depressing aggregate demand. This contractionary effect may be more than offset by a rise in export volume, if η is sufficiently large, generating an expansionary real devaluation.

Taylor's condition is modified when we allow mark-ups to be affected bythe level of investment, as e.g. in Ball (1964), Harcourt (1972), Eichner (1973), Harris (1974), Wood (1975) and Harcourt and Kenyon (1976). The general idea is that investment generates both a quantity and a price effect. A rise in investment leads to a greater capacity utilisation and to a larger *net* profit margin, for a given mark-up. This is the quantity effect. But those authors note that the higher investment may be financed by a higher mark-up rate, increasing the *gross* profit margin as well. The rise in the mark-up implies a higher price level, which may offset the increase in capacity utilisation and in the net profit margin. If the rise in the mark-up is too high, capacity utilisation will actually be reduced. This indicates the possibility of a *negative* multiplier effect of a higher investment demand, pointed out by Harcourt (1972) (see also Harris 1974): this is what we call the *Harcourt effect*. In an open economy, this effect can be reinforced or counterbalanced depending on whether real appreciation—following a rise in the domestic price level—is contractionary or expansionary.

A rise in the mark-up following increased investment expenditure has a double impact in an open economy: first, as in the closed economy case, it reduces the real-wage rate; second, in the absence of a nominal exchange-rate devaluation, it reduces the degree of international competitiveness of the economy, thus causing a *real exchange-rate appreciation*.

In order to present the argument more clearly, let us assume that investment demand has an autonomous component (g_0) and that the mark-up rate is an increasing function of the level of *autonomous* investment: changes in entrepreneurs' 'animal spirits', which shift the investment demand schedule, are those that influence pricing decisions. Formally:

$$g^i(i, u) = g_0 + g_1(i, u) \tag{13}$$

$$p = \psi(g_0)(wb + ea),\ \partial p/\partial g_0 = p_y\psi_g > 0 \tag{14}$$

Harcourt (1972) notes that the positive relationship between ψ and g_0 may produce a negative multiplier effect of investment on output. There are two channels through which this relationship may cause contractionary effects: a reduction in the real wage ω or in the real exchange rate θ, following an increase in g_0. An increase in the price level reduces ω ($\bullet\omega'\bullet p = \omega_p < 0$) and θ ($\bullet\theta'\bullet p = \theta_P < 0$). Therefore:

$$\omega = \omega(p(\psi(g_0)));\ \partial\omega/\partial g_0 = \omega_p p_\psi \psi_g < 0 \tag{15a}$$

$$\theta = \theta(p(\psi(g_0)));\ \partial\theta/\partial g_0 = \theta_p p_\psi \psi_g < 0 \tag{15b}$$

Let us now reconsider the product market equilibrium condition (11) with the new version of the investment function, Equation (13):

$$g_0 + g_1(i, u) + \gamma - \sigma(\omega, \theta)u + x(\theta, u^*) = 0 \tag{11'}$$

Taking into account the direct immediate impact of a higher g_0 on capacity utilisation ($\bullet u/\bullet g_0 = u_g > 0$), the resulting multiplier effect will be:

$$\frac{du}{dg_0} = \frac{1 - (\sigma_\omega\omega_p p_\psi\psi_g + \sigma_\theta\theta_p p_\psi\psi_g)u - \sigma(\psi,\theta)u_g + x_\theta\theta_p p_\psi\psi_g}{\sigma(\omega, \theta) - g^i_u} \tag{16}$$

As we presuppose the denominator to be positive, a negative multiplier obtains when the numerator is negative.

It may be demonstrated that a negative multiplier is produced when the condition below holds:

$$\begin{matrix}\theta_p p_\psi \psi_g[\eta(x/\theta) - \sigma_\theta u] < \sigma_\omega\omega_p p_\psi\psi_g u + \sigma(\omega, \theta)u_g - 1 \\ <0 \quad >0 \text{ or } <0 \quad >0 \qquad\qquad >0\end{matrix} \tag{17}$$

The term in brackets will be negative or positive depending on whether condition (12) for contractionary devaluation holds. Expression (17) therefore provides an important link between Taylor's condition for contractionary devaluation and the 'Harcourt effect' in an open economy.

When Taylor's condition (12) is satisfied [$\eta(x/\theta) - \sigma_0\theta < 0$], the likelihood of the 'Harcourt effect' (a negative multiplier) is reduced, as the LHS of (17) is

positive. Otherwise, a sufficient condition for the 'Harcourt effect' to arise is that the RHS of (17) is non-negative. The possibility of a negative multiplier increases if Taylor's condition (12) is *not* met.

REAL-WAGE RESISTANCE

In the previous sections, potential sources of class conflict are disguised under the form of 'supply elements' (the mark-up, the money wage, the exchange rate, etc.). Furthermore, social conflict itself is 'frozen' during each short period as a result of fixed money wage and mark-up (with the exception of the brief consideration of the possibility of the 'Harcourt effect').

The hypothesis of rigid money wages is widespread in Cambridge distri bution models. It is usually associated with real-wage flexibility, so that changes in the price level can eliminate imbalances between planned saving and investment. Marglin (1984) notes that although fixed money wages *may* be acceptable in a short-period analysis, it is misleading in a long-period context:

> Needless to say, this assumption is implicit rather than explicit in the neo-Keynesian [or Cambridge] model. The rhetoric is, after all, Marxian: capitalists' and workers' interests are firmly opposed to one another. But in the neo-Keynesian model, the essential Marxian notion of class *conflict* gives way to a kind of functionalism. Class interests may be opposed, but the class struggle is long since over. Victorious, the capitalist class works its will on a supine working class. Whatever the money wage might be, capitalists are free to utilise the central bank and the banking system to manipulate money, credit, and prices to their own ends while workers look on helplessly.
>
> (Marglin 1984:475)

It does not follow that the Cambridge theory is incompatible with flexible *money* wages. Marglin's (1984) own model suggests that the Cambridge theory may be reconciled with conflict-inflation theory, which points to the attempts on the part of workers to maintain a certain target real wage, within a long-period framework. As to the short period, Harcourt's (1965) two-sector (bread and steel) model allows for the endogenous determination of money wages. In the bread (or consumption goods) sector, money wages are a function of the price of bread, technical progress and unemployment level; in the steel (or capital goods) sector, money wages are set in accordance with the current employment requirement in that sector and with the money wage established in the bread sector. While Harcourt's (1965) model permits the endogenous determination of money-wage *levels,* Solow and Stiglitz (1968) show that the Kaldorian distribution formula may be valid in the short period even when there is a

process of ongoing inflation, with an endogenously determined *rate of change* of money wages.

Earlier attempts to incorporate class conflict and real-wage resistance into the Cambridge theory of distribution include Joan Robinson's concept of the inflation barrier. This barrier establishes the minimum real wage workers are prepared to accept, and saving may be prevented from adjusting to an increased level of investment at full capacity through changes in the price level, if the actual real wage falls below the minimum acceptable level, generating pressures for money-wage increases.

The necessity of taking into account some form of real-wage resistance is even greater in an open economy, when exchange-rate depreciations or rises in the prices of imported goods reduce the real wage for any given money-wage rate, imposing a decline in workers' welfare standards. Mainstream open economy macroeconomics has considered the effects of real-wage resistance under fixed exchange rates (e.g. Dornbusch 1980) and flexible exchange rates (e.g. Sachs 1980).

Real-wage resistance may be incorporated in our framework in a straightforward manner. Suppose that wage earners' require a minimum real-wage rate ω_0 determined in accordance with historically acquired consumption standards. If $\omega<\omega_0$, they will demand money-wage increases which will produce a wage-price spiral which can only be terminated if the minimum real wage is restored or if the bargaining power of trade unions is dampened. Money wages are presupposed to follow a simple dynamic rule:

$$\frac{dw/dt}{w(t)} = f[\omega_0 - \omega(\theta)], f' > 0, f(0) = 0, f(-) = 0, \quad \frac{d\omega}{d\theta} < 0 \tag{18}$$

The rate of growth of money wages is an increasing function of the *aspiration gap* (Rowthorn, 1977) given in brackets and the actual real wage is a decreasing function of the real exchange rate. If $\omega=\omega_0$, no changes in money wages occur. In contrast with Sach's (1980) analysis, if $\omega>\omega_0$, money wages will not be reduced: an asymmetry between inflation and deflation is assumed, so that money wages are still rigid downwards. In the short period, we suppose that the target real wage is independent of the unemployment rate.

Given the technical coefficients a and b and the 'degree of monopoly' factor ψ, there is only one real exchange rate θ_0 compatible with the minimum real wage acceptable by labour (Dornbusch, 1980:70 – 71). The *maximum feasible* exchange rate is given by:

$$\theta_0 = \frac{p^*}{a}\left(\frac{1}{\psi} - \omega_0 b\right) \tag{19}$$

which follows from the price equation (4a). Equation (19) shows that the greater is the minimum acceptable real wage, the lower is the maximum feasible real

exchange rate, given *a, b, p** and ψ. The real exchange rate can be depreciated up to θ_0—beyond this point, it hits the inflation barrier.

Associated with the maximum feasible real exchange rate, there is a maximum feasible share of profits π, π_{max}, which may be well below unity. Let us assume that real depreciation is expansionary, that is, Taylor's condition (12) is not met. The share of profits is then an unambiguously increasing function of the real exchange rate. Greater competitiveness has a *double* positive impact on *X:* it reduces the share of wages in unit prime costs and increases the level of capacity utilisation. In a fixed-exchange rate regime, however, π_{max} can be increased by expenditure-risen policies that donot require devaluation. Therefore, the limit to π imposed by θ_0 is operative only insofar as the fiscal and monetary policy parameters are kept unchanged.

The case of a flexible nominal exchange rate is slightly more complicated. With a credit-driven money supply, as seen in the previous section above, a fiscal expansion or a reduction in the domestic interest rate will be reinforced by an endogenous nominal and real exchange-rate depreciation. Expenditure-increasing policies, *ceteris paribus,* increase both the share of profits and the real exchange rate. Since π and θ increase together (probably, although, by different rates), if the new degree of competitiveness exceeds θ_0, we may say that the corresponding share of profits *overshoots* its new maximum feasible level. Real-wage resistance and the ensuing wage-price spiral cause the real exchange rate to appreciate, partially crowding out net exports until $\theta=\theta_0$. At this level of the real exchange rate, the share of profits is reduced to its new maximum feasible level.

The analysis of the open-economy case with real-wage resistance thus shows that the real exchange rate is yet another element of the macroeconomy, which can drive it towards the inflation barrier.

CONCLUDING REMARKS

In this paper we developed a short-period version of the so-called Cambridge theory of distribution for a small, open economy.

The basic results of the previous sections can be summarised as follows:

1 Effective demand affects the distribution of income in the short period, even if profit margins are fixed, provided that there exists *overhead labour* in the economy;
2 the share of profits in the value of output is determined by 'supply-side' and 'demand-side' elements: the long-period versions of the Cambridge theory of distribution tend to concentrate on the latter;
3 the effects of exchange-rate fluctuations on the share of profits are, in principle, ambiguous, as both the 'supply side' (via pricing decisions) and the 'demand side' (via capacity utilisation) are affected by the exchange rate;

4 Mundell's (1961) model is shown to be a simple but illuminating vehicle for analysing exchange-rate flexibility with a credit-driven, demand-determined money supply;
5 the 'Harcourt effect' of a negative investment multiplier is more likely to occur when exchange-rate depreciation is expansionary;
6 real-wage resistance can be easily incorporated in the basic structure by using a simple formulation of the 'aspiration gap' encountered in conflict-inflation models.

Needless to say, those conclusions would be affected if some of our assumptions were relaxed, allowing, for instance, for stock-flow dynamics with asset accumulation going on, alternative expectations-formation rules, or multi-sectoral settings. More contemporary analyses of money-wage determination could also be contemplated, along the lines suggested by Solow(1992).

ACKNOWLEDGEMENTS

I would like to thank G.C.Harcourt, K.J.Coutts, M.Baranzini and Jean-Luc Rosinger for extremely helpful comments on an earlier version of this paper, presented at the XXII National Economics Conference, December 1994, Florianópolis, Brazil.

NOTES

1 The findings, interpretations and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the World Bank, its executive directors, or the countries they represent.
2 Such as Atsumi (1959 – 60), Sen (1963), Harcourt (1963, 1965, 1972), Solow (1967), Solow and Stiglitz (1968), Riach (1969), Harris (1974), Asimakopulos (1975) and Rimmer (1993).
3 Understood as changes in the money rate of interest determined by the central bank.
4 With a perfectly demand-determined supply of money, the so-called 'compensation thesis' applies. The essential idea of this thesis is that any *undesired* money supply expansion generated from foreign sources (payments received and capital inflows denominated in foreign currency) is *compensated* for by a corresponding reduction in domestic base money, as the public uses the inflow of foreign currency to decrease outstanding debts with the commercial banks and the latter use the extra funds to reduce their debts with the central bank (Lavoie 1992: ch. 5).
5 In order that the covered interest parity condition is satisfied.
6 Mundell (1961:513): 'Monetary policy is in reality extremely complicated with many side effects, depending on how it is implemented. I shall assume that it takes the form of a reduction (increase) in interest rates, and that its main impact is to increase investment and reduce capital imports'.

7 We could also add a sixteenth equation for a sixteenth variable, total labour force *(L_f),* such that $L_f=L_{f0}$, i.e. a given labour force in the short period. We assume that $L^d<L_f$.

8 Alternatively, we may consider a 'normalised' rate of capacity utilisation, dividing the 'true' rate *u* by the capital-output ratio *k* in order to obtain *u* above.

9 *I/K* is taken to be a decreasing function of *i* and an increasing function of the *expected* rate of profits r^e: $I/K=I/K\ (i, r^e)$. Assuming, in a Keynesian fashion, that current events shape long-term expectations, r^e is supposed to be equal to the *current* average rate of profits ($r^e=r$). With a given mark-up on average prime costs, *r* is an *index of capacity utilisation* ($r=f(u), f>0$). Therefore, we mayinterpret *I/K* as an increasing function of the rate of capacity utilisation *(u),* as a proxy for the expected rate of profits, which resembles the one adopted by Dutt and Amadeo (1993).

10 A similar specification is adopted in the short-period models of Meade and Vines (1988) and Sarantis (1990 – 1, 1993). See also Mundell (1961).

11 This conventional formulation resembles the Mundell-Fleming original approach and once again follows Meade and Vines (1988) and Sarantis (1990 – 1, 1993). Our departure from the Mundell-Fleming model can be seen in our hypotheses regarding pricing, investment decisions, income distribution and its effects on saving, and money supply endogeneity.

12 See Taylor (1991:145): 'a higher import cost acts as a vehicle for foreign saving'.

13 In the present model, a nominal exchange rate devaluation, with fixed money wages, implies a real exchange-rate devaluation, as the price level changes less than proportionately after a rise in *e*.

REFERENCES

Araujo, J.T. (1994) 'Four essays in the Cambridge theory of distribution and growth', unpublished Ph.D. dissertation, University of Cambridge.

——(1995) 'Kaldor's Neo-Pasinetti model and the Cambridge theory of distribution', *The Manchester School,* vol. LXIII, no. 3, September: 311 – 17.

Arestis, P. and Eichner, A.S. (1988) 'The post-Keynesian and institutionalist theory of money and credit', *Journal of Economic Issues,* XXII, 4, December: 1003 – 21.

Asimakopulos, A. (1969) 'A Robinsonian growth model in one sector notation', *Australian Economic Papers,* 8:41 – 58.

——(1970) 'A Robinsonian growth model in one sector notation: an amendment', *Australian Economic Papers,* 9:171 – 6.

——(1975) 'A Kaleckian theory of income distribution', *Canadian Journal of Economics,* 8 (3), August: 313 – 33.

Atsumi, H. (1959 – 60) 'Mr Kaldor's theory of income distribution', *Review of Economic Studies,* XXVII:108 – 18.

Ball, R.J. (1964) *Inflation and the Theory of Money,* London: Allen and Unwin.

Bhaduri, A. and Marglin, S.J. (1990) 'Unemployment and the real wage: the economic basis for contesting political ideologies', *Cambridge Journal of Economics,* 14 (4), December: 375 – 93.

Blecker, R.A. (1989) 'International competition, income distribution and economic growth', *Cambridge Journal of Economics,* 13 (3), September: 395 – 412.

Dornbusch, R. (1980) *Open Economy Macroeconomics,* New York: Basic Books.
Dutt, A.K. and Amadeo, E.J. (1993) 'A post-Keynesian theory of growth, interest and money', in M.Baranzini and G.C.Harcourt (eds) *The Dynamics of the Wealth of Nations: Growth, Distribution and Structural Change. Essays in Honour of Luigi Pasinetti,* New York: St Martin's Press, 181 – 205.
Eichner, A.S. (1973) 'A theory of the determination of the mark-up under oligopoly', *Economic Journal,* 83, 332:1184 – 200.
Hahn, F.H. (1951) 'The share of wages in national income', *Oxford Economic Papers,* June: 147 – 57.
Harcourt, G.C. (1963) 'A critique of Mr Kaldor's model of income distribution and economic growth, *Australian Economic Papers,* 2:20 – 36.
——(1965) 'A two-sector model of the distribution of income and the level of employment in the short run', *Economic Record,* XLI: 103 – 17.
——(1972) *Some Cambridge Controversies in the Theory of Capital,* Cambridge: Cambridge University Press.
Harcourt, G.C. and Kenyon, P. (1976) 'Pricing and the investment decision', *Kyklos,* 29, fasc. 3:449 – 77.
Harris, D.J. (1974) 'The price policy of firms, the level of employment and distribution of income in the short run', *Australian Economic Papers,* 13, 22:144 – 51.
Krugman, P. and Taylor, L. (1978) 'Contractionary effects of devaluation', *Journal of International Economics,* 8:445 – 56.
Lavoie, M. (1992) *Foundations of Post-Keynesian Economic Analysis,* Aldershot: Edward Elgar.
Marglin, S.J. (1984) *Growth, Distribution and Prices,* Cambridge MA: Harvard University Press.
Meade, J.E. and Vines, D.A. (1988) 'Monetary policy and fiscal policy: impact effects with a New Keynesian "assignment" of weapon to targets', in S.Howson (ed.) *The Collected Papers of James Meade,* London: Unwin Hyman.
Moore, B. (1988) *Horizontalists and Verticalists: The Macroeconomics of Credit Money,* Cambridge: Cambridge University Press.
Morishima, M. (1992) *Capital and Credit: A New Formulation of General Equilibrium Theory,* Cambridge: Cambridge University Press.
Mundell, R. (1961) 'Flexible exchange rates and employment policy', *Canadian Journal of Economics and Political Science,* XXVII, 4, November: 509 – 17.
Riach, P.A. (1969) 'A framework for macro-distribution analysis', *Kyklos,* 22: 542 – 65.
Rimmer, R. (1993) *Income Distribution in a Corporate Economy,* Aldershot: Edward Elgar.
Robinson, J. (1956) *The Accumulation of Capital,* London: Macmillan.
Rowthorn, R.E. (1977) 'Conflict, inflation and money', *Cambridge Journal of Economics,* 1:215 – 39.
——(1981) 'Demand, real wages and economic growth', *Thames Papers in Political Economy,* London: Thames Polytechnic.
Sachs, J. (1980) 'Wages, flexible exchange rates and economic policy', *Quarterly Journal of Economics,* 94:731 – 47.
Sarantis, N. (1990 – 1) 'Distribution and terms of trade dynamics, inflation and growth', *Journal of Post-Keynesian Economics,* 13:175 – 98.
——(1993) 'Distribution, aggregate demand and unemployment in OECD countries', *Economic Journal,* 103, March: 459 – 67.

Sen, A.K. (1963) 'Neo-classical and neo-Keynesian theories of distribution', *Economic Record,* 39:53 – 64.

Solow, R.M. (1967) 'Distribution in the long and short run', in J.Marchal and B. Ducros (eds) *The Distribution of National Income,* proceedings of a conference held by the International Economic Association, London: Macmillan.

——(1992) 'Hahn on the share of wages in national income', in P.Dasgupta, D.Gale and E.Maskin (eds) *Economic Analysis of Markets and Games: Essays in Honour of Frank Hahn,* Cambridge MA: MIT Press.

Solow, R.M. and Stiglitz, J.E. (1968) 'Output, employment, and real wages in the short run', *Quarterly Journal of Economics,* LXXXII:537 – 60.

Taylor, L. (1991) *Income Distribution, Inflation and Growth: Lectures on Structuralist Macroeconomic Theory,* Cambridge MA: MIT Press.

Weale, M., Blake, A., Christodoulakis, N., Meade, J.E. and Vines, D. (1989) *Macroeconomic Policy,* London: Unwin Hyman.

Wood, A. (1975) *A Theory of Profits,* Cambridge: Cambridge University Press.

19
THE POLITICAL ECONOMY OF MACROECONOMIC POLICY

Pulapre Balakrishnan[1]

INTRODUCTION

Writing soon after the event, Paul Krugman accounted for the Mexican currency crisis of early 1995 as an instance of how the economic opinion of the day interacting with the workings of financial markets can at times generate a bubble. The economic opinion then current has been referred to as the 'Washington consensus'. *Washington* was meant to capture not just the United States government,

> but all those institutions and networks of opinion leaders centred in the world's de facto capital—the International Monetary Fund, World Bank, think tanks, politically sophisticated investment bankers, and worldly finance ministers, all those who meet each other in Washington and collectively define the wisdom of the moment.
>
> (Krugman 1995:28 – 9)

The 'consensus' was originally named[2] by John Williamson, then of the Washington-based Institute of International Economics, to include ten different aspects of economic policy. However, Krugman has summarised it aptly:

> It is the belief that Victorian virtue in economic policy—free markets and sound money—is the key to economic development. Liberalise trade, privatise state enterprises, balance the budget, peg the exchange rate, and one will have laid the foundations for an economic take-off.
>
> (Krugman 1995:29)

While, in this context, 'free markets' refer to the elimination of barriers to international trade, 'sound money' has generally been interpreted to be the pursuit of policies that are believed to contribute to zero inflation.

The Washington consensus has had a major influence on economic policy the world over. A vehicle on which these ideas have ridden the crest of the tide of

economic ideas is the policy-related lending of the Washington-based multilateral financial institutions. Principally, when countries constrained by balance of payments have approached these institutions they have been required to undertake 'fiscal correction', defined as a reduction of the public sector deficit. A smaller deficit, it has been argued, lowers the rate of inflation and increases the rate of growth of output. Graft on the idea that external imbalances are caused by excessive pubic expenditure, and the deficit watchers have an additional argument to focus on the deficit.

This paper is an analysis of the possibilities associated with the use of the fiscal deficit as an instrument of macroeconomic policy. It appears important enough to do this, even after the somewhat diminished prestige of the said consensus following upon the fiasco in Mexico, an economy once hailed as an exemplar, in early 1995. This is so because economic ideas are perhaps more prone to hysteresis effects than an economy itself! Therefore I first consider, in the first section, the use of the fiscal deficit as an instrument, and point to some likely pitfalls in its use as such. Next, in the second section, I discuss the determination of the fiscal deficit, which process can crucially affect the outcome of 'fiscal correction'.

THE FISCAL DEFICIT AS AN INSTRUMENT

Traditionally, a macroeconomic stabilisation programme in an open economy has been seen as having three objectives.[3] These are the lowering of the inflation rate, the attainment of a sustainable current-account deficit and the reduction of the level of the external debt. More recently, the criterion that stabilisation programmes must be growth-oriented has been proposed (Dornbusch 1990). While it has not been made clear what this entails precisely, it may reasonably be interpreted as suggesting that stabilisation programmes at least address the question of growth. Thus it has increasingly come to be understood that a credible stabilisation programme must aim to create the conditions for recovery and thus growth. However, it is with regard to the attainment of the first two objectives mentioned above that the fiscal deficit has been proposed as an instrument; in fact, the principal one.

Reining-in external imbalance: 'the twin deficits' view

The idea that the fiscal deficit may be used as an instrument to tackle the current-account deficit has now reached centre stage of the mainstreamdiscussions on stabilisation policy. However, it is not based on a behavioural model, nor is it even an empirical regularity. It is based overwhelmingly on the US experience of the 1980s when supply-side economics, rising defence outlays and restrictive monetary policy, all together, contributed to mounting internal and external deficits. The alleged origins of an all-too-neat story of the twin deficits are, it

would appear, in the standard national-income identity expressing the equality between aggregate supply and aggregate demand in equilibrium. Recall that

$$Y = C + I + G + X - M$$

can be re-arranged, since $Y=C+S+T$, to yield[4]

$$(M-X) = (I-S) + (G-T)$$

Certainly, this expression sustains the interpretation that the current-account deficit and the government's deficit, the same as the fiscal deficit, are linked so long as private-sector behaviour, reflected in the magnitude of the excess or shortfall of private investment over private savings, remains constant. As with all identities, however, causality is a different matter, but this has not deterred the voicing of the claim that the fiscal deficit *causes* an external one.

Recently, Feldstein (1992) has raised some simple but fundamental questions about the efficacy of cutting the budget deficit in an attempt to reduce the current-account imbalance. Feldstein's contention is that the crucial long-run macroeconomic relation is that between saving and investment. Shifts in the budget deficit may be interpreted as shifts in national savings. Now there is no reason why the adjustment must be made via changes in the current-account balance alone. Feldstein's view is that it could equally be made via changes in investment. On the question of the nature of the adjustment, in particular the extent of the improvement in the trade deficit, it is observed that this must depend upon the interest elasticity of investment, the extent to which the exchange rate is affected and the extent to which the trade deficit responds to the exchange rate. This alternative view of the likely consequence of budget cuts can be seen from a re-arrangement of the original national-income identity to read:

$$(S-D) = I + NX$$

where $(S-D)$ is national savings, S being the savings of the private sector and D being the public-sector deficit, I, as before, is private investment, and NX is net exports, defined here as the current account balance. Feldstein also cites the cases of economies as diverse as the UK and Mexico, where success in reducing the budget deficit actually has been accompanied byincreasing current account deficits. More egregiously, when in early 1995 the Mexican economy faced a balance of payments crisis, the public sector budget was in surplus (Sachs *et al.* 1995).

Can it be said that the proponents of budget-deficit reduction as a means to reducing the external deficit might actually have a more sophisticated mechanism in mind? Specifically, it may be argued that budgetary deficits lead to inflation which contributes to exchange-rate overvaluation and a natural emergence of a current-account deficit. However, if the avoidance of overvaluation of the exchange rate is the issue, it can be taken care of by allowing the exchange rate to float freely. My reason for rejecting this version, however, is based on a belief

that the said proponents actually have in mind a more direct relation between the two deficits.

The control of inflation

When demand management is the prescription for inflation control, and it mostly is, the fiscal deficit is yet again the principal instrument it seems, though tight monetary policy is often recommended in support of fiscal contraction here. Where inflation is an aggregate excess-demand phenomenon, this is, of course, appropriate. Moreover, there is no denying the role of demand factors in most inflations. However, the widespread phenomenon of stagflation must raise the distinct possibility of cost-push often being at work as part of the inflationary process. This would immediately suggest the lack of efficacy of pure demand-based strategies, such as demand management. Non-linearities no doubt exist, and there might thus come a stage when contraction has gone sufficiently far that even demand-management policies begin to curb inflation.[5] Even though the policy-maker implementing such a strategy might get to resemble Napoleon marching on, through the icy Russian winter, only to find Moscow burning![6] I am, of course, referring to the concomitant output loss. Demand management that promises 'disinflation' (the scaling down of prices, quantities unchanged) begs the question.

That fiscal restraint is a necessary and sufficient criterion for the lowering of the inflation rate meets with little challenge today. However, it was not always so. The view under consideration follows from the assumptions that the fiscal deficit adds to aggregate demand and that inflation is a symptom of excess aggregate demand. On the other hand, it has long been recognised that inflation could well occur in an economy with excess capacity, at least in some sectors, and no excess demand in the aggregate. This had stemmed from an appreciation of the importance of price-formation mechanisms for the inflationary process (Balakrishnan 1991).

The question of the behaviour of prices cannot be divorced from that of how they are formed. An approach to analysing price behaviour, and the nature of an inflationary process, is provided by the Hicksian dichotomy of'flexprice' and 'fixprice' markets (Hicks 1973). In the first instance, the characterisation pertains to the process of formation of prices. In the flexprice markets, prices clear any demand-supply imbalance. Naturally, these prices are flexible. The price in the many markets for agricultural commodities is seen to approximate such behaviour. In the fixprice markets, on the other hand, firms fix prices in relation to costs. The level of the mark-up is assumed to remain invariant with respect to demand. This implies quantity adjustment as the norm. Price in the market for manufactures is seen to approximate such behaviour. It is important to note that this does not imply that demand does not affect price whatsoever, only that it must affect prices via any impact that it might have on costs. For, since

price changes are cost-determined, prices are changed only .when costs alter. This characterisation of price behaviour according to areas of economic activity is originally due to Kalecki (Kalecki 1954).

It must by now be apparent that the nature of an inflationary process is intimately linked to the mechanisms of price formation. These mechanisms impinge upon the efficacy of alternative anti-inflationary strategies. At least, from the Hicksian scheme, it is clear at the very outset that demand management, originally an expression referring to the fine-tuning of output, can have only a limited application,[7] because demand affects price behaviour only in certain sectors of the economy.

Would granting, for the sake of argument, that inflation is an excess demand phenomenon render the fiscal deficit the appropriate instrument? No. Because increases (decreases) in the fiscal deficit can be combined with decreases (increases) in fresh purchasing power injected into the system. The additional purchasing power enters the system as a part of the fiscal deficit is monetised. It is interesting that till the late 1980s, in the discussion of the public finances of India, 'budget deficit' referred to the overall budgetary deficit of the Union government inclusive of public borrowing. It is easily seen that it is this definition of the deficit that unambiguously points to a change in the purchasing power brought about by government. By 'purchasing power' is, of course, meant the injection of high-powered money into the system. The extent to which changes in the stock of money are reflected in a change in aggregate demand is a separate issue. It arises because of the possibility of contemporaneous changes in velocity offsetting changes in the stock of money, thus leaving aggregate spending power unchanged. However, it is of the essence.

An argument relating inflation to 'fiscal' deficits, albeit in the long run, does, however, exist. This is the 'unpleasant monetarist arithmetic' (Sargent and Wallace 1982) which points to a higher inflation rate when deficits are debt-financed as opposed to money-financed. The intuition is that in the long run, when the debt is repaid by way of printing money, the money stock would have gone up by principal *plus* interest which, obviously, is more than would be the case had money been printed to start with. Thisimmediately makes all deficits matter for inflation, except that its empirical relevance depends upon the maturity structure of debt.

Votaries of traditional stabilisation programmes argue that a stable macroeconomic environment defined as emerging out of fiscal correction, defined as reduction of the fiscal deficit, is necessary for growth to resume. Extremists would hold that it is a sufficient condition. However, quite to the contrary, indiscriminate deficit reduction too could have some unintended consequences. I now turn to this aspect.

SOME UNINTENDED CONSEQUENCES OF DEFICIT REDUCTION

The experience of many developing economies that were subjected to conventional macroeconomic stabilisation programmes during the 1980s is that it is not automatic that these programmes lead to a resumption of growth, let alone ensuring the transition to a higher growth path. Thus discussion has shifted to the policies that lead an economy from stabilisation to growth. This has also been influenced by a development in economic theory. The short-run focus of macroeconomic policy has for too long been considered to be a legitimate delimitation. This was so perhaps because macroeconomic theory itself was considered to be concerned with the short run. It was a theory of employment. A separate theory, 'growth', dealt with the long term. It took the recognition of the importance of 'hysteresis' in the working of economies to bridge the gap between the analyses of economic activity in the short-run and the long-run, respectively. Hysteresis is the property of dynamic systems that the stationary equilibrium is a function of the initial conditions and/or the trajectory to the steady state (Cross 1988). Now, traditional macro-policy that focuses on the short run could actually end up as 'short termist'. A prime example comes from the theory of employment where it has been argued that government policy can actually shift the 'natural' rate. For instance, if the economy is kept depressed for long, it is conceivable that the 'natural rate' might rise. Closer to the concerns of this paper, consider the possibility that dealing with an allegedly 'overheated' economy might actually affect its long-term trajectory, that is, growth. It has always been recognised that conventional programmes for the stabilisation of an economy can be damaging to growth. This has been the experience of many developing economies. It has by now been identified that this occurs via contraction of investment, particularly private investment. Investment is often influenced directly by the stabilisation policy pursued and indirectly due to the incentive structure that comes into the picture by the very adoption of a stabilisation programme. I consider the direct effects first.

Demand management and private investment

Contraction has been the *sine qua non* of demand management during a conventional stabilisation programme. This can affect investment. Efforts to reduce aggregate demand inevitably reduce output. If the accelerator mechanism prevails, and it does widely, investment must decline. Any initial downturn in output could also affect investment via expectation of recovery. Investment is delayed until the recovery occurs. This is often especially so in projects with a short gestation and has the effect of delaying the recovery further.

As to why demand management affects output, consider the manner in which the principal instruments work. First, monetary policy is considered to affect aggregate demand via investment, either through interest-rate policy or by credit

rationing.[8] Second, where 'crowding in' rather than 'crowding out' characterises the relation between public and private investment, the fiscal deficit, which would determine the extent of capital spending by government, would have a direct bearing on private investment. Now, when demand management is intended to be deflationary, and fiscal and monetary policy is geared towards this objective, it can directly reduce private investment and thus output. Serven and Solimano (1993) cite studies that point to a complementarity between private and public investment in developing economies. Heilbroner (1992) points to work on the US economy that reveals much higher multiplier effects of public as compared to private investment. This has a bearing on the practice of cutting public expenditure to contain the fiscal deficit. The political economy of expenditure cuts in the face of resistance from interest groups often implies that it is the public investment programme that is the first to go. This, of course, is no longer pure economics. Politics enters in a definite way.

Private investment and the incentive structure

Recently, there has been some interesting work on the behaviour of investment during stabilisation programmes. Most contributions play on the old idea of investment being an irreversible decision having to be taken in an uncertain environment. In fact, the importance of uncertainty flows directly from the irreversibility of the investment expenditure which takes on the form of a sunk cost (Pindyck 1991). It is pointed out that the opportunity cost of investment is the 'option to wait'.

It is argued that the uncertainty involved makes the 'macroeconomic environment' to be as important as tax incentives or the interest rate in determining investment. The literature, however, is somewhat silent on what is meant by 'the macroeconomic environment', even though inflation is cited. A related argument stresses the relationship between 'credibility' of government policy and the incentive structure faced by firms. The argumentis made more strongly in the instance of structural reforms, but it does hold for stabilisation programmes too. Quite simply, if the private sector finds the government's stated intentions (either to reform or to stabilise the economy) less than credible it is likely to postpone investment decisions for the irreversible nature of investment decisions makes for the possibility of an irreversible mistake. Finally, an aspect that has a bearing on the incentive structure faced by potential investors is not related to either stabilisation policies or to structural reforms, but to a characteristic of most economies that undertake either. This is debt overhang. The repayment of debt involves transfer to the country's creditors, which leads to expectations of large swings in taxes or in aggregate demand in the future. Essentially, the repayment of debt is equivalent to a tax on incomes in the future, and is believed to act as a dampener on private investment.

While the incidence of each of the above mechanisms is hard to establish, the question of the behaviour of investment in general and of private investment in particular tends to get ignored in assessments of the progress of stabilisation programmes. To the extent that capital formation is vital to growth, this is a major oversight. Proponents of the conventional package will argue that stabilisation of the economy will bring about an increase in investment by itself. It is the existence of such an automatic mechanism that is in question.

ON THE DETERMINATION OF THE FISCAL DEFICIT

It is obvious that if the fiscal deficit is to be used as a policy instrument it must be subject to the policy-maker's control. There are, however, two sets of arguments which point out that this may be a little more easily assumed than it is actualised. The first has to do with the fact that a large part of government expenditure and revenue depends upon the state of the economy. What the government can of course do is to set the rates of tax and expenditure, such as an *ad valorem* subsidy. The volumes are determined by the level of income. For instance, in an economy where agriculture accounts for a large part of national income and has large indirect effects on non-agricultural production, a substantial part of income determination is not amenable to control, in the short run, either. The idea that the budget deficit is partly endogenous has, interestingly, always been recognised by elementary macroeconomic textbooks. This is precisely the implication of the idea of an 'automatic' or 'built-in' fiscal stabiliser. The idea is that, with tax rates given, tax revenues decline when the economy is in a downswing. Now the budget deficit automatically increases, thus contributing an expansionary stimulus. The second set of arguments which point to a difficulty in a government's controlling the fiscal deficit is motivated by considerations ofpolitical economy. These recognise the susceptibility of government budgetary decisions to lobbying by special-interest groups. Outcomes are directly affected by concessions obtained and by benefits in the form of directed expenditure that are wrung out of a government. Predatory special-interest groups can effectively frustrate efforts to reduce the deficit during stabilisation episodes.

There exist various theories, starting with one by Schumpeter, of why government deficits arise in the first place.

The weakness of the 'tax state' as visualised by Schumpeter is prominent among the political-economy explanations. Taylor (1990) however, provides some purely economic explanations, which are interesting for their ability to turn conventional arguments regarding the behaviour of macroeconomic variables on their head. Consider first an external shock which might be brought about through either an adverse shift in the terms-of-trade or a rise in the rate of interest on external debt. Such shocks are contractionary. Where governments try to offset them by expansionary fiscal policy, any existing fiscal deficit expands. A second situation arises in the context of a government taking over the external

obligations of an economy, as has commonly occurred in Latin America. Now the economy has to run a current-account surplus to finance the debt payments, while internally the state must mobilise resources to obtain the foreign exchange to make these payments. There is, as it were, a 'double transfer' problem. The government can tax exporters or raise taxes more generally. Or it can borrow in the domestic capital markets, thus contributing to the fiscal deficit. Note that these two examples serve the purpose of reversing the causality implicit in the twin deficits argument discussed in the first section above. Some further illuminating examples are provided by Rakshit (1991).

Reducing the fiscal deficit: some pitfalls

The foregoing discussion enables us to understand the record with regard to attempted fiscal correction during stabilisation programmes. The first is that a reduction in the fiscal deficit is not so easily brought about. In an accounting sense, an inability to reduce the deficit reflects the state's inability, or unwillingness, to raise additional resources while at the same time lowering its expenditure. However, this is not as interesting an observation as is the second of the two mentioned above, which is that when fiscal deficits are reduced, this is brought about disproportionately via capital expenditure cuts. The consequences of capital expenditure cuts are not discussed in this paper. Here, only the reasons why this pattern recurs are considered.

In a heterogeneous democratic polity, budgetary outcomes are seldom technocratic choices made by Olympian bureaucrats, but rather the result of interest-group pressures and the desire of the political class to nurture theirconstituencies. It is now easy to see that when governments contract with international lending agencies to reduce the fiscal deficit they find it easiest to cut capital expenditure. Current expenditure affects immediately, and in an obvious way, the standard of living of some groups, and hence are likely to be resisted by those powerful enough to do so. On the other hand, the effects of capital spending are long-drawn-out, or at least imagined to be. Essentially, future generations are easily bypassed. In terms made popular by Hirschman (1981), since they have no 'voice' future generations cannot hold out the threat of 'exit' from the polity either. It is for this reason that fiscal 'correction' is often accompanied by a disproportionate reduction of capital as opposed to current expenditures on government account.

CONCLUSION

Surprisingly, at a time when a somewhat fundamentalist interpretation of economic theory has undermined the prestige of macroeconomic analysis, macroeconomic policy is a major preoccupation of governments. In fact, in the argument that all that governments need to do is to maintain *fiscal discipline* and

a neutral exchange rate, is the suggestion that macro policy is all that matters. The evidence from across the world is that the reality of the matter is more complex.[9] In particular, we understand, first, that there are pitfalls in the way of 'fiscal correction' defined in terms as simple as a reduction of the fiscal deficit, and, second, that economic recovery is not automatic even once fiscal adjustment is undertaken. The macroeconomic environment, and thus macroeconomic policy, has a clearly specified but only limited role. Policy regimes are enabling rather than causative, and the mainsprings of sustained economic growth are clearly more than just sound money and free markets.

The deeper question is how certain opinions capture the high ground of economic policy debates, and thus affect the lives of millions. Paul Krugman has provided an interesting account:

> endless rounds of meetings, speeches, and exchanges of communiqués… occupy much of the time of economic opinion leaders. Such interlocking social groupings tend at any given time to converge on a conventional wisdom, about economics among other things. People believe certain stories because everyone important tells them. Indeed when a conventional wisdom is at its fullest strength, one's agreement with that conventional wisdom becomes almost a litmus test of one's suitability to be taken seriously.
>
> (Krugman 1995:36)

ACKNOWLEDGEMENTS

I was encouraged by Jayati Ghosh to write this paper for Geoff, and by Arunava Sen to think critically about the role of macroeconomic policy in general. While I gratefully acknowledge these acts, all responsibility is mine.

NOTES

1 The findings, interpretations and conclusions expressed in this paper are entirely those of the author. They do not represent the views of the World Bank, its Executive Directors, or the countries they represent.

2 Williamson 1990:7.

3 What the objectives are is not just an academic issue. It is important that one assesses a stabilisation programme on its own terms, or to be precise, in terms of the stated objectives of its votaries. In this connection, it is interesting to note that the objectives as stated above are those recognised quite widely See Khan (1987) and IMF (1987).

4 In the above, Y is GNP, C is private consumption, I is private investment, S is private savings, G is government expenditure, T stands for taxes, while M and X are imports and exports, respectively, both inclusive of invisibles.

5 For the price of agricultural goods may be expected to be related to industrial activity, and the industrial wage must begin to respond to the level of unemployment.
6 To borrow from Amartya Sen's conceptualisation of the dilemma that was then being faced by the Thatcherite policy machine in early 1980s Britain.
7 With hindsight we might now say that the disenchantment, itself due more to developments in economic theory, with fine-tuning of output was too easily interpreted as evidence in favour of the possibility of fine-tuning the price level. Instances of the failure of demand management in the economies of Europe can be explained in terms of the history of the individual situation, most notably economies acting on their own as opposed to acting in concerted fashion. More generally, they are to be interpreted as saying something about the relative efficacy of 'micro' versus 'macro' intervention. And this is as much relevant to price policy as it is to stabilisation policy.
8 The first of course assumes that investment is interest-elastic. That credit rationing is likely to curtail output in general and investment in particular, is not controversial. Those interested in the development of macroeconomics would notice the model of monetary policy that underlies each of these accounts. It differs from the classic statement of Friedman (1968) which posits a monetary policy affecting output only via its influence on the labour-supply decision.
9 For international evidence across type of economy see Taylor (1991). More recently, we have a World Bank study of the relationship between public-sector deficits and macroeconomic performance in nine developing economies, which reports the finding that the correlation between the deficit and any one variable such as inflation, the interest rate or the exchange rate, 'is near zero' (Easterly *et al.* 1994).

REFERENCES

Balakrishnan, P. (1991) *Pricing and Inflation in India,* Delhi: Oxford University Press.

Cross, R. (ed.) (1988) *Unemployment, Hysteresis, and the Natural Rate Hypothesis,* Oxford: Blackwell.

Dornbusch, R. (1990) 'Policies to move from stabilisation to growth', *Proceedings of the First World Bank Conference on Development Economics,* Washington DC: World Bank.

Easterly, W, Rodriguez, C.A. and Schmidt-Hebbel, K. (eds) (1994) *Public Sector Deficits and Macroeconomic Performance,* New York: Oxford University Press for the World Bank.

Feldstein, M. (1992) 'The budget and trade deficits aren't really twins', *Challenge,* 35: 60 – 3.

Friedman, M. (1968) 'The role of monetary policy', *American Economic Review,* 58: 1 – 17.

Heilbroner, R. (1992) 'The deficit: a way out', *New York Review of Books,* November.

Hicks, J. (1973) *The Crisis in Keynesian Economics,* Oxford: Blackwell.

Hirschman, I. (1981) 'Exit, voice, and the state', in *Essays in Trespassing: Economics to Politics and Beyond,* Cambridge: Cambridge University Press.

International Monetary Fund (1987) 'Theoretical aspects of the design of Fund-supported adjustment programmes', Occasional Paper no. 55.

Kalecki, M. (1954) 'Costs and prices', reprinted in *Selected Essays on the Dynamics of Capitalist Economies,* Cambridge: Cambridge University Press.

Khan, M.S. (1987) 'Macroeconomic adjustment in developing economies: a policy perspective', *The World Bank Research Observer,* 2:23 – 42.

Krugman, P. (1995) 'Dutch tulips and emerging markets', *Foreign Affairs,* Jul/Aug, vol. 74:28 – 45.

Pindyck, R. (1991) 'Irreversibility, uncertainty and investment', *Journal of Economic Literature,* 29:1140 – 8.

Rakshit, M. (1991) 'The macroeconomic adjustment programme: a critique', *Economic and Political Weekly,* 26:1977 – 88.

Sachs, J., Tornell, A. and Velasco, A. (1995) 'The collapse of the Mexican peso: what have we learned?', Working Paper no. 5742, Cambridge MA: National Bureau of Economic Research.

Sargent, T.J. and Wallace, N. (1982) 'Some unpleasant monetarist arithmetic', *Federal Reserve Bank of Minnesota Quarterly Review,* 5:1 – 17.

Serven, L. and Solimano, A. (1993) 'Debt crisis, adjustment policies and capital formation in developing countries: where do we stand?', *World Development,* 21: 127 – 40.

Taylor, L. (1990) 'Fiscal issues in macroeconomic stabilisation', *Ricerche Economiche,* 44:213 – 30.

——(1991) *Varieties of Stabilisation Experience: Toward Sensible Macroeconomics in the Third World,* Oxford: Clarendon Press.

Williamson, J. (1990) 'What Washington means by policy reform', in J.Williamson (ed.) *Latin American Adjustment: How much has happened?,* Washington DC: Institute for International Economics.

20
UNEMPLOYMENT HYSTERESIS IN AN OPEN ECONOMY

Terry O'Shaughnessy

INTRODUCTION

Geoff Harcourt combines a vigorous commitment to the Keynesian tradition of policy-orientated macroeconomic analysis with a lifelong concern to sort out the puzzle of the relationship between the short-run performance of an economy and its long-run prospects. In this, he is part of a tradition of post-Keynesian analysis which began with Joan Robinson, Richard Kahn and Nicholas Kaldor. Just as they were, Geoff is very willing to confront the orthodoxies of the time. This chapter, which begins with an account of a current orthodoxy—the 'natural' or 'non-accelerating inflation' rate of unemployment—and explores possible links between short-run outcomes and long-run constraints on economic performance, is motivated by these concerns of his.

The main focus of the chapter is on the notion of 'hysteresis'. Economies—like other complex systems—display 'hysteresis' when their current underlying, structural properties depend on the past history of the system. In the first section, the case of hysteresis in the natural rate of unemployment is considered. The second section discusses two mechanisms through which hysteresis in unemployment might occur. The first of these mechanisms operates through the labour market; the second is via changes in the productive capacity of the economy. The next section introduces the distinction between linear and non-linear systems, and attempts to clarify competing usages of the terms 'persistence' and 'hysteresis' in the description of such systems. In the fourth section a small linear model of an open economy is constructed in which current levels of unemployment influence the underlying 'natural' rate. The implication is that policy—say, to deflate the economy in order to reduce inflationary pressure—will have long-run effects.

THE NATURAL RATE OF UNEMPLOYMENT AND HYSTERESIS

The dominant theory now used to explain unemployment and inflation in market economies involves the specification of an expectations-augmented Phillips curve.[1] The key, organising concept in this theory is the 'equilibrium', 'natural' or 'non-accelerating inflation' rate of unemployment. As Phelps (1967:255) put it, the equilibrium unemployment rate is that 'at which the actual rate of inflation equals the expected rate of inflation so that expected inflation remains unchanged'. The interaction of the Phillips curve mechanism with agents' expectations means that it is possible to use expansionary fiscal or monetary policy in order to reduce unemployment at the cost of higher inflation, but that this reduction in unemployment will be achieved only for so long as expectations take to adjust. Advocates of this theory are thus rather sceptical about the usefulness of demand-orientated policies to reduce unemployment. Conversely, it is possible to reduce inflation at the cost of higher unemployment. However, this cost will be only temporary, since eventually expectations of inflation will adjust downwards, allowing unemployment to fall back to its natural rate. Advocates of this theory are therefore cautiously optimistic about the usefulness of deflationary policies to reduce inflation, while acknowledging the temporary cost. They note that these costs can be reduced if the natural rate is lowered, or if expectations can be made to adjust more quickly. They tend to favour supply-side policies designed to reduce the natural rate of unemployment and institutional changes (such as introducing coordinated wage setting or adopting a fixed exchange rate with a low inflation currency) designed to reduce inflationary expectations.

A difficulty with the theory is that the key concept of the 'natural' rate of unemployment' or NAIRU has proved rather elusive. Empirical estimates seem very sensitive to the model of wage- and price setting employed and the mechanism for determining expectations (Setterfield *et al.* 1992). Moreover, the natural rate seems itself to be a function of past values of the actual rate of unemployment (Jenkinson 1987, 1988; Blanchard and Summers 1986). This phenomenon—known as 'hysteresis'—has the effect of undermining the very notion of a 'natural' rate of unemployment. It also calls into question the policy implications of the theory. For instance, if the cost of deflationary policy is a *permanently* higher natural rate of unemployment, such a policy looks much less attractive. Conversely, if, by reducing the actual level of unemployment, an expansionary policy can help to reduce the natural rate of unemployment, it becomes a much more defensible approach to adopt.

MECHANISMS

Broadly speaking, there are two mechanisms that may be invoked to account for hysteresis effects in the unemployment-inflation interaction. The first focuses on

the wage bargain. According to this view, the deflationary impact of restrictive policies is mainly via wage-setting behaviour. Higher levels of unemployment discourage inflationary wage settlements, but only if the unemployed can exert downward pressure on such settlements by being potential alternative employees. If high unemployment has persisted for some time, some of the long-term unemployed will be 'outsiders' as far as wage setting is concerned. Any slight tightening of the labour market can then be exploited by the 'insiders' who will be able to secure higher wage settlements from their employers as a result. Persistent unemployment, by creating more 'outsiders' and fewer 'insiders', raises the point at which inflationary pressures are likely to set in; in other words, it raises the natural rate of unemployment. While sharing a similar focus on wage-setting behaviour, some investigators stress 'duration' effects (Hargreaves-Heap 1980; Clark and Summers 1982; Nickell 1987; Arestis and Skott 1993) while others stress 'membership' effects (Blanchard and Summers 1987; Lindbeck and Snower 1987).

The second mechanism focuses on the productive capacity of the economy. According to this view, an important effect of deflationary policy is the accelerated scrapping of existing capacity. Firms are also discouraged from investing in new capacity. As productive capacity adjusts downwards to the actual level of capacity that is being utilised, the inflationary potential of any subsequent increase in demand becomes more acute. There are a number of ways in which such inflationary pressures can be transmitted. Most obviously, capacity-constrained firms may raise their prices in response to higher demand. Second, employees of such firms may be able to secure higher wages than otherwise, since their firms will be both more willing and more able to agree to higher settlements. A third channel may also be called into play in the tradable-goods sector, where demand that cannot be satisfied by capacity-constrained firms will be met by higher imports. This will moderate price and wage rises in capacity-constrained firms in this sector, but at the cost of creating pressure on the exchange rate. If this pressure leads to a fall in the value of the currency there will eventually be a rise in the general level of tradable-goods prices, which will be transmitted sooner or later to wages and to non-traded goods prices.

It is difficult to separate labour market and capacity effects in studies of hysteresis,[2] though there is a consensus, at least as far as the United Kingdom is concerned, that insider-outsider effects are rather more important than capacity constraints; see, for example, Layard and Nickell (1987), Layard and Bean (1989), Bean and Gavosto (1990). My own view is that labour market constraints have been overestimated and capacity constraintshave been underestimated in recent discussions of hysteresis, though it is hard to assign precise weights to each, given the data problems involved.[3]

HYSTERESIS OR PERSISTENCE?

In addition to disagreements about the source of hysteresis effects, there are also disagreements about how the dynamics should be specified when hysteresis effects are incorporated into economic models. This has given rise to a number of distinctions, including that between hysteresis and *persistence* and between *full* (or *pure*) and *partial* hysteresis. Unfortunately, these distinctions are not used consistently in the economics literature.

An early (and still common) approach defines hysteresis in terms of linear systems of equations with unit roots (in the case of difference equations) or zero roots (in the case of differential equations). Applied to unemployment hysteresis, this would mean that current unemployment depends on past values of unemployment with coefficients that sum to 1. This is the view taken by Blanchard and Summers (1986:17, n1), although they are prepared to use the term 'hysteresis' more loosely to include cases where the sum of the coefficients is close to but not necessarily equal to 1. Others, such as Layard *et al.* (1991), distinguish between pure hysteresis (where the coefficients sum to 1) and partial hysteresis (where the sum is less than 1), while others still (Franz 1987) make the same distinction using the terms hysteresis and persistence.

An important property of linear systems with unit roots is that they 'remember' all the shocks to which they are subject. However, this is not a property of hysteresis in physical systems where 'memory' is more selective. Analysis of such non-linear systems (Krasnosel'skii and Pokrovskii 1989; Mayergoyz 1991) has developed techniques for modelling processes in which the current state of a system depends on *non-dominated extremum values* of past shocks. A number of economists (Cross 1993, 1994, 1995a, 1995b;[4] Archibald 1995; Amable *et al.* 1995) have argued that economic systems with hysteresis should be modelled in a similar way. They would prefer to confine the term 'hysteresis' to such non-linear systems and describe linear systems with unit (or near unit) roots as displaying 'persistence'. This approach has the advantage that it brings usage into conformity with that in the physical sciences. It also directs attention to the fact that the microfoundations of economic systems displaying hysteresis are almost always highly non-linear, and that this has important implications for the analysis of economic aggregates. In the past macroeconomists have assumed (or hoped) that aggregation would 'smooth out' the effects of non-linearities at the micro level. This will not occur in systems with hysteresis. The challenge is to develop techniques and data sources that allow analysis to move from the micro to the macro level in a coherent way.[5] In the meantime, we should beaware that studying models with unit roots (as in the next section) provides some insight into the persistent effect of shocks (or policy choices) on economic systems.

HYSTERESIS IN AN OPEN ECONOMY

In a number of important papers in the early 1980s, Buiter and Miller (1981a, 1981b, 1982, 1983) studied the role of real exchange-rate overshooting during deflationary episodes. They argued that current (or anticipated) monetary tightening will cause the exchange rate to appreciate. The consequent loss of competitiveness, by reducing output below its capacity level, is one of the channels through which deflationary monetary policy operates in reducing domestic inflation. In their model, output eventually returns to its capacity level, but in the meantime there will be a—perhaps substantial—loss of output. In their 1983 paper, Buiter and Miller's main concern was to investigate what determined the magnitude of this output loss, and to explore whether alternative policies can achieve a given desired reduction in inflation at a smaller cost. They found that, when the policy choice is confined to conventional fiscal and monetary instruments, the time integral of the output loss is invariant and depends only on the slope of the Phillips curve and the parameter which determines how quickly inflationary expectations adapt.[6] To reduce the output cost associated with a given reduction in inflation below this level it would be necessary to deploy other policies, such as cutting indirect taxes or instituting an incomes policy, with a view to having a direct impact on 'core' or 'expected' inflation.[7]

Throughout their analyses, Buiter and Miller assume that the economy eventually returns to its full capacity output. In other words, the output loss is temporary, while the reduction in inflation is permanent. This means that reasonably long-sighted agents who care both about output and inflation should be prepared to put up with such losses, while seeking to reduce them if possible. However, a more worrying possibility is that the output loss may prove very persistent—or even permanent—because of the phenomenon of hysteresis.

THE MODEL

The model, which is set out in Equations (1) – (7), is based on a simplified version of Buiter and Miller (1983), with the addition of a hysteresis effect, as captured by Equation (7). All variables, except for r, r^* and π are in logs.

$$\frac{dp(t)}{dt} = \theta_1(u^*(t) - u(t)) + \pi(t) \tag{1}$$

$$\frac{d\pi(t)}{dt} = \theta_2 \left(\frac{dp(t)}{dt} - \pi(t) \right) \tag{2}$$

$$m(t) = ky(t) - \lambda r(t) + p(t) \tag{3}$$

$$y(t) - y^*(t) = -\gamma\left(r(t) - \frac{dp(t)}{dt} + \delta(e(t) + p^*(t) - p(t))\right) \tag{4}$$

$$\begin{aligned} y(t) - y^*(t) &= -\theta_5(u(t) - u^*(t)) \\ \tilde{y}(t) &= y^*(t) + \theta_5 u^*(t) \end{aligned} \tag{5}$$

$$\frac{de(t)}{dt} = r(t) - r^*(t) \tag{6}$$

$$\frac{du^*(t)}{dt} = \theta_3(u(t) - u^*(t)) \tag{7}$$

The variables are defined as follows: *p(t)* the domestic price level; thus *dp(t)/dt* is the rate of inflation; *u(t)* the unemployment rate; *u*(t)* the natural rate of unemployment; *π(t)* expected inflation; *m(t)* the nominal stock of money; *y(t)* output; *y*(t)* capacity output, corresponding to *u(t)=u*(t); y• (t)* full employment output, corresponding to *u(t)*=0; *r(t)* the nominal interest rate; *r*(t)* the foreign nominal interest rate; *e(t)* the exchange rate (the domestic currency price of foreign currency).

Equation (1) is the short-run, expectations-augmented Phillips curve. It describes the outcome of a wage- and price-setting process in which agents are influenced by the level of activity relative to productive capacity, and by their expectations of inflation. Equation (2) specifies how wage- and price-setters adjust their expectations about inflation in the light of any discrepancy between the actual inflation rate and the rate that they previously expected.[8] Equation (3) is the LM curve; it based on an exogenously set supply of money, *m(t),* and a demand function for real money balances which depends positively on output and negatively on the nominal interestrate. Equation (4) is the IS curve; it says that the demand for domestic output depends negatively on the real interest rate and on the foreign currency price of domestic goods. Equation (5) is Okun's law, relating changes in output to corresponding changes in employment. Equation (6) is the uncovered interest rate parity condition. It is based on the fact that capital is perfectly mobile at a world interest rate, *r(t)* (for convenience, the world interest rate is taken as constant). Note that, as in Dornbusch's model (1976) and subsequent contributions (including Buiter and Miller's) based on it, it is assumed that agents operating on the foreign exchange market form their expectations rationally or (in non-stochastic settings, like this one) with perfect foresight. This is in contrast to wage- and price-setters who, according to Equation (2), form their expectations adaptively. Finally, Equation (7) specifies how quickly the natural rate of unemployment adjusts if the actual rate of unemployment is above (or below) the current natural rate; in other words, this is the hysteresis effect.

INITIAL CONDITIONS

Before examining the dynamics of this system, it necessary to specify some initial conditions and to clarify the distinction between what Buiter and Miller describe as 'predetermined' and 'jump' variables. Suppose that we begin from a situation at t=0 in which inflation, inflationary expectations and the rate of growth of the money supply are all equal at some initial (high) level, μ_0.

$$\frac{dp(0)}{dt} = \pi(0) = \frac{dm(0)}{dt} = \mu_0 \tag{8}$$

Suppose, also, that this situation has persisted for some time, so that the actual rates of output and unemployment equal the corresponding natural rates.

$$u(0) = u^*(0);\ y(0) = y^*(0) \tag{9}$$

A policy to reduce inflation may now be implemented by a money growth rule by which $m(t)$ grows at a new, lower, rate, μ.

$$m(t) = m_0 + \mu t;\ t > 0 \tag{10}$$

Here m_0 is the stock of money inherited by the system at t=0. At t=0 the system also inherits the current values of u^* (0), $p(0)$ and $\pi(0)$ (u^*_0 p_0 and π_0 =μ_0], respectively but $e(t)$ jumps to a new value (say, e_0), as does $u(t)$. Thus$e(t)$ and $u(t)$ are 'jump' variables, while $u^*(t)$, $\pi(t)$ and $p(t)$ are 'predetermined' in Buiter and Miller's sense.

The system (1) – (10) may be solved by taking Laplace transforms, solving the set of simultaneous equations that results and then taking inverse Laplace transforms.[9] This method yields a complete characterisation of the time path of $u(t)$, $y(t)$, $p(t)$, $e(t)$ and so on. However, we can immediately determine the level of unemployment at the end of this process (which is also equal to the new value of the NRU/NAIRU) by applying the final value theorem. This yields

$$u(\infty) = u^*(\infty) = u_0^* + \frac{\theta_3}{\theta_1\theta_2}[\mu_0 - \mu] \tag{11}$$

Note that the permanent increase in unemployment that results from a reduction in monetary growth (and in inflation) from μ_0 to μ is proportional to this change in the rate of growth of the money supply, and depends positively on θ_3 (the coefficient in the hysteresis equation) and negatively on θ_1 (the slope of the short-run Phillips curve) and on θ_2 (the coefficient in the expectations equation which determines how quickly expectations adapt to discrepancies between expected and actual inflation). The long-run increase in unemployment is independent of the coefficients in the IS and LM equations and specifically of δ which relates output to competitiveness and is thus a measure of the economy's openness. In fact, Equation (11) gives the same value for this increase in

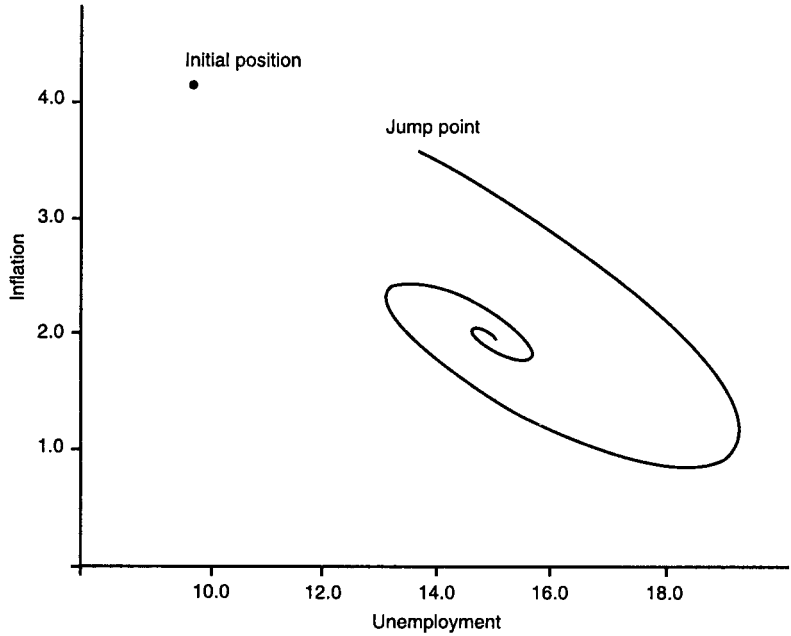

Figure 20.1 Path traced out by the system in the traditional Phillips curve space, following a reduction in the rate of growth of the money supply

unemployment as that obtained in an earlier study (O'Shaughnessy 1993) of hysteresis in a simpler closed economy model. Of course, the time path of $u(t)$ does depend on δ, as does that of $e(t)$. In particular, the extent of the jump in the exchange rate at $t=0$ depends on δ.

Insight into the dynamics of the system may be obtained from Figures 20.1 and 20.2, which are drawn in the traditional Phillips curve space, with unemployment on the horizontal axis and inflation on the vertical axis. Initially, the system is at the point (10, 4). Following the policy change it moves immediately to the point labelled 'jump point' in Figure 20.1 and then evolves along a spiral-shaped path to a new, long-run position. The point to which the system jumps initially is on the short-run Phillips curve (Equation (I)) while the point at which it ends up lies on a locus defined by Equation (II). Figure 20.2 shows both these loci; the former is labelled ab and the latter ac. Figure 20.2 also shows the role that the degree of openness of the economy (captured by the parameter δ in Equation (4)) plays in determining the path the system follows. The more open the economy (the larger is δ), the larger is the initial jump along the short-run Phillips curve. However, the point at which the system ends up is determined by the locus ac and the new value of μ (the rate of growth of the money supply); it is independent of δ.[10]

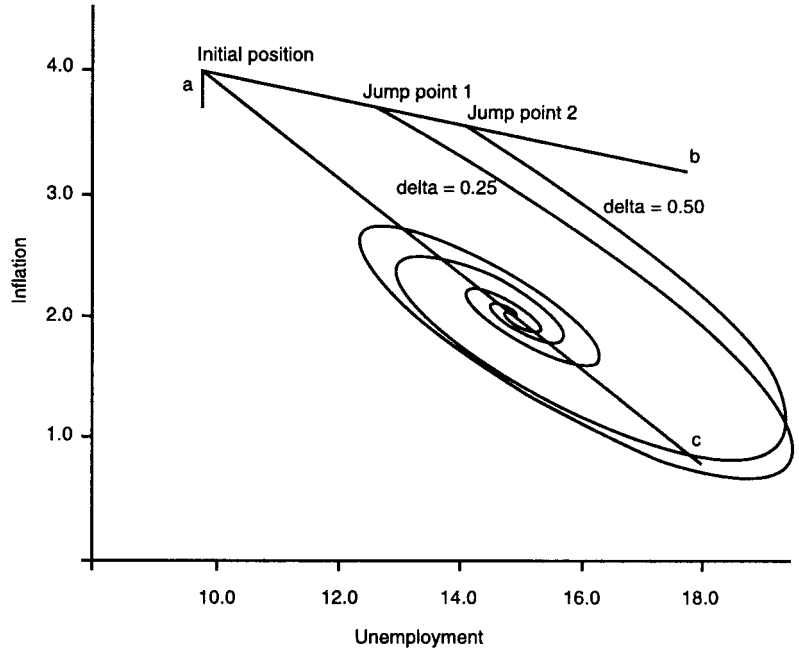

Figure 20.2 Two paths traced out by the system, corresponding to different values of δ

Finally, Figures 20.3 and 20.4 offer another way of viewing the behaviour of the system in which the focus is on the current account of the balance of payments. Deflationary policies of the type analysed in this chapter have both short-run and long-run effects on the current account. In the short run, the fall in domestic activity will tend to move the current account into surplus as the demand for imports falls. Moreover, if trade volumes take some time to adjust following the jump in the exchange rate, a (reverse) 'J-curve' effect will temporarily improve the trade account. On the other hand, the loss in competitiveness will eventually lead to a fall in export volumes and a shift by domestic purchasers away from locally produced tradables and towards imported substitutes.

In the longer run there will be scrapping of capacity in the tradable goods sector of the economy. This means that, other things being equal,[11] the level of unemployment that is consistent with current account balance will be higher. If we identify this 'Current Account Balance Rate of Unemployment' (or CABRU) with u^*, we can write

$$ca(t) = \theta_6(u(t) - u^*(t)) + \theta_7(e(t) - p(t)) \qquad (12)$$

where *ca(t)* is the current account of the balance of payments[12] and θ_6 and θ_7 are suitable constants. A useful way of capturing the behaviour of the current account

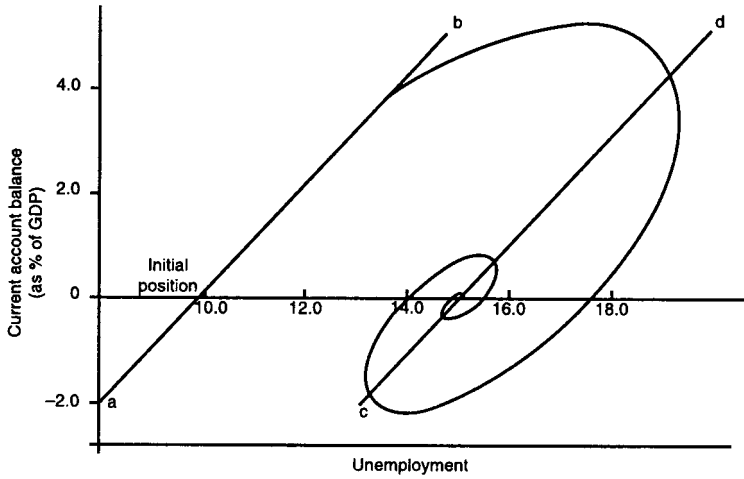

Figure 20.3 Current account of the balance of payments plotted against unemployment, following a reduction in the rate of growth of the money supply

which brings out both the short-run and long-run processes identified here is to plot the current account against unemployment, as is done in Figure 20.3 (for the model) and Figure 20.4 (for the UK economy for the period 1958:1 – 95:1, using quarterly data). In the short run there appears to be a tradeoff between the current account and the level of unemployment: this is given by the line ab in Figure 20.3 and by the dotted lines labelled 1958 – 66, 1974 – 80 and 1984 – 8 in Figure 20.4. In fact, in the absence of hysteresis effects, Equation (12) predicts that the data points in Figure 20.4 will lie in a more or less narrow band around the line ab. The slope of this line is determined by the parameter θ_6; its width is determined by θ_7 and by the other model parameters that determine the range of values that *e(t)—p(t)* can assume. If hysteresis effects are present, the position of this short-run tradeoff will shift, say to cd.

The historical data used to plot Figure 20.4 show both effects. During certain periods most of the data points lie on or near the three dotted lines shown in the Figure; in fact, these lines are OLS regression lines drawn using data from the indicated sub-periods. However, there has been a series of shifts in the position of this short-run trade off. Between 1958 and 1966 the position of this line was such that current account balance appeared to be consistent with an unemployment level somewhat under 500,000. By 1974 – 80 the tradeoff had shifted outwards so that unemployment of about 1.3 million was consistent with current account balance.[13] However, a more dramatic shift took place during the first part of the 1980s, when the terms of the tradeoff between unemployment and the current account worsened to such an extent that unemployment of over 3 million seemed to be required for current account balance. In each case, these shifts were

associated with changes in the capacity of the tradable goods sector of the UK economy, especially in manufacturing.[14]

Figure 20.4 also illustrates more recent developments in the UK economy, during which this process began to operate in the opposite direction. One of the effects of the boom in the late 1980s was to reduce unemployment and worsen the current account; this corresponds to a move down along the regression line labelled 1984 – 8 in Figure 20.4. But another effect was to begin shifting this tradeoff inwards. Inspection of data points from the period 1989:4 to 1991:2 suggests that, by then, the current account balance rate of unemployment had fallen to about 2.5 million, though subsequently the tradeoff shifted out again as policy in the UK was tightened in order to maintain membership of the Exchange Rate Mechanism. Finally, the tradeoff shifted in again following the UK's exit from the ERM. Note that the inward shift in the unemployment-current account tradeoff during the boom in the late 1980s was associated with a sharp rise in investment in the manufacturing sector[15] which led to some increase in capacity, though not sufficient to make up for capacity scrapped in earlier periods.

Thus it appears that the UK's experience is consistent with hysteresis effects of the type captured in the model described in this chapter. However, it would be expecting too much of the model for it to describe precisely the dynamics linking unemployment, inflation, competitiveness and the balance of payments. For a start, the treatment of expectations makes too stark a distinction between the way wage- and price-setters behave and the way those who trade on the foreign exchange market behave. Wage- and price-setters are assumed to form their expectations adaptively, while in the foreign exchange market expectations are rational. In practice, there is bound to be a whole range of ways in which participants in various marketsform their expectations, with neither of these extreme cases being a fully satisfactory description. Second, the hysteresis effect, which links changes in the natural rate of unemployment to the extent of the discrepancy between the actual rate of unemployment and the natural rate, is formulated in a very simple way which neglects non-linear, 'selective memory' effects. Third, if, as is suggested in this chapter, the natural rate of unemployment in an open economy is related to capacity in the tradable goods sector, then all the factors which affect the latter will play a role in determining the former. In particular, it may well be the case that discrepancies between the actual and the natural rates of unemployment have different impacts, depending on whether the discrepancy is positive or negative and on whether the discrepancy is expected to be temporary or to persist. Finally, the treatment of the current account is rudimentary, since there is no place for the effect of lags, while interest, profits, dividends and transfers are not modelled.

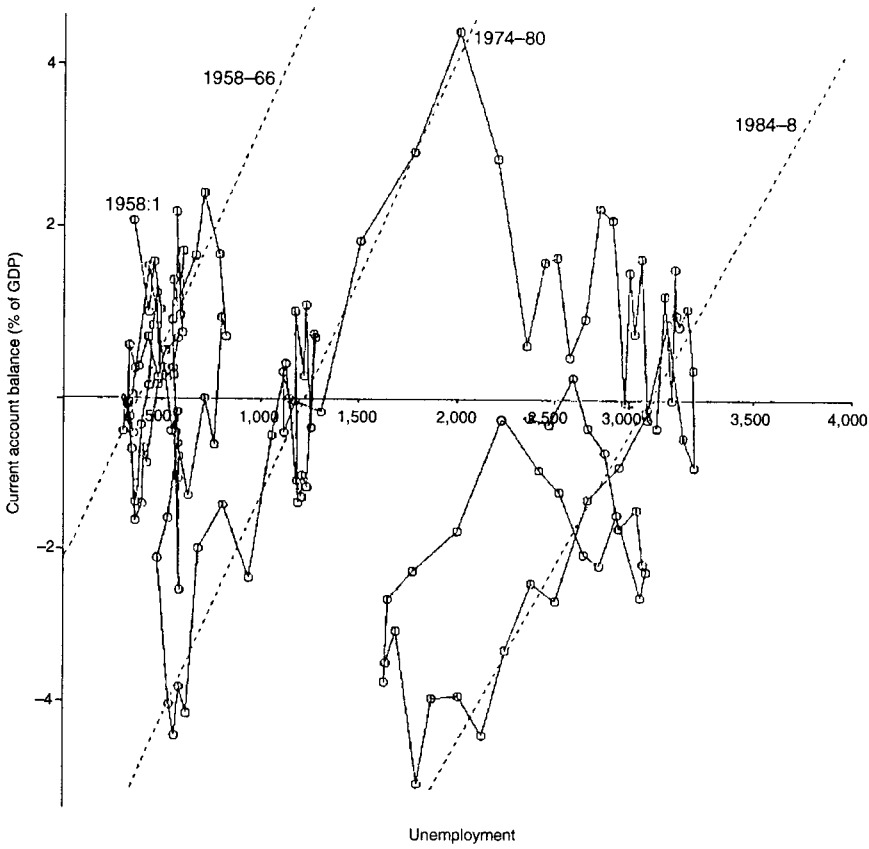

Figure 20.4 Current account of the balance of payments plotted against unemployment ('000)

Source: United Kingdom quarterly data, 1958Q1 – 1995Q1.

CONCLUSION

This chapter explores the link between exchange-rate overshooting during deflationary episodes and changes in the NAIRU, especially changes in the NAIRU that arise from changes in the productive capacity of an economy's tradable goods sector. Such hysteresis effects give rise to a long-run tradeoff between inflation and unemployment.[16] The slope of this relationship depends on the slope of the short-run Phillips curve and the coefficients which specify how fast inflationary expectations adjust to changes in inflation and how fast capacity adjusts to changes in capacity utilisation. However, it does not depend on how open the economy is—in other words, on how much real exchange-rate

overshooting takes place—although the dynamics of the adjustment process certainly do depend on this factor.

The policy implications of the analysis put forward in this chapter are disturbing. If hysteresis effects are significant and the effects of deflationary policies are persistent, it follows that the costs of reducing inflation are much higher than analyses which neglect hysteresis would suggest. This might encourage some policy-makers to support reflationary programmes with a view to getting the hysteresis process to work in reverse. A more long-sighted approach would involve measures to accelerate the adjustment of inflationary expectations during deflationary episodes (and decelerate the process during reflationary ones). It would also involve policies designed to delay the scrapping of capacity during recessions and to encourage investment in new capacity without relying entirely on the (inevitably short-lived) stimulus of the next reflationary boom.

NOTES

1 See Phelps (1967, 1970) and Friedman (1968). A large number of empirical studies use this framework; see, for example, Layard and Nickell (1986) and *Layard et al.* (1991).

2 One reason is that firms often scrap equipment and lay off workers as the result of a single decision, so there is a close match between the number of 'insiders' and remaining productive capacity. In addition, data on capacity utilisation are limited in coverage and difficult to interpret.

3 Soskice and Carlin (1989) and Rowthorn (1995) present models in which capital constraints play an important role. Empirical studies of capacity, investment and scrapping in the UK include Driver (1986), Driver and Morton (1991), Ingham *et al.* (1988), Price *et al.* (1990), Mayes and Young (1994), Price (1995).

4 Note that in his 1987 paper, Cross analyses hysteresis in terms of a continuous-time linear model with a zero root.

5 For an interesting exploration of these issues, see Göcke (1994).

6 In the notation of Buiter and Miller (1983), the output loss equals the reduction in inflation multiplied by the factor $1/\xi$ in the notation used in this chapter this factor is $\theta_5/\theta_2\theta_1$.

7 If the authorities announce a lower rate of monetary growth and this is believed, expected inflation may respond immediately, with a similar effect; see Buiter and Miller (1981a).

8 Note that Buiter and Miller (1983) describe π in their version of the expectations-augmented Phillips curve equation as 'trend' or 'core' inflation, and they identify it with the (right-hand) time derivative of the money supply. This is equivalent to assuming that agents determine their expectations of future inflation by observing the current rate of growth of the money supply. While there may appear to be some justification for this, in the sense that, in both Buiter and Miller's model and the model described in this chapter, the *long-run* outcome is that the inflation rate equals the rate of growth of the money supply, it is not clear why agents should

form their expectations in this way in the short run. For this reason I prefer the interpretation and specification of $\pi(t)$ given in the text.

9 Details of the solution, together with a worked numerical example, are in O'Shaughnessy (1994).

10 In the study already cited (O'Shaughnessy 1993) of hysteresis in a model of a closed economy, it was shown that the vigour with which a deflationary policy was pursued (measured by how quickly the rate of monetary growth is adjusted downwards) affects the path the economy follows but not its end point. In an open economy the parameter δ plays a similar role, capturing, as it does, the 'vigour' with which the real exchange-rate overshooting effect impacts the deflationary process.

11 In this context, wage rates and rates of return on capital in the tradable and non-tradable goods sectors of the economy.

12 Conveniently expressed as a percentage of GDP. This equation captures the activity and competitiveness effects on the current account, but it does not attempt to model the lags between a change in competitiveness and changes in import and export volumes that produce J-curve effects.

13 In fact, this shift seems to have occurred in two steps. Data points from the first half of the 1970s appear to lie along a line (not drawn in Figure 20.4) which passes through 750,000 on the horizontal axis. There is then a further shift outwards (associated with the first oil shock) in 1973 – 4.

14 Since the 1973 oil price shock (which was adverse for the UK) is clearly visible in Figure 20.4, readers may wonder why the 1979 oil price shock is not. (This shock was favourable to the UK, which was by then a significant oil exporter.) The answer lies in the fact that, for a period in the 1970s, shrinking capacity in manufacturing was matched fairly closely by increasing energy exports, so that, in Figure 20.4, data points from the period 1974 to 1980 lie along the same line. This phenomenon lent support to 'Dutch disease' explanations of events during this period, though such explanations were less plausible later, when the energy sector ceased to expand but manufacturing capacity continued to shrink.

15 Gross fixed investment in manufacturing industry was 63 per cent higher in 1988 – 9 than it had been in 1981 – 3.

16 This is the locus defined by Equation (11) above.

REFERENCES

Amable, B., Henry, I, Lordon, F. and Topol, R. (1995) 'Hysteresis revisited: a methodological approach', in R.Cross (ed.) (1995c) *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis,* Cambridge: Cambridge University Press.

Archibald, G.C. (1995) 'Hysteresis and memory in the labour market', in R.Cross (ed.) (1995c) *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis,* Cambridge: Cambridge University Press.

Arestis, P. and Skott, P. (1993) 'Conflict, wage determination and hysteresis in UK wage determination', *Journal of Post-Keynesian Economics,* 15, 365 – 86.

Bean, C.R. and Gavosto, A. (1990) 'Outsiders, capacity shortages and unemployment in the United Kingdom', in J.H.Dreze and C.R.Bean (eds) *Europe's Unemployment Problem,* Cambridge MA: MIT Press.

Blanchard, O.J. and Summers, L.H. (1986) 'Hysteresis and the European unemployment problem', in S.Fischer (ed.) *NBER Macroeconomics Annual 1986,* Cambridge MA: MIT Press.

——(1987) 'Hysteresis in unemployment', *European Economic Review* 31, 288 – 95.

Buiter, W.H. and Miller, M. (1981a) 'Monetary policy and international competitiveness', *Oxford Economic Papers,* 33, 143 – 75.

——(1981b) 'The Thatcher experiment: an interim report', *Brookings Papers on Economic Activity* 2, 315 – 67.

——(1982) 'Real exchange rate overshooting and the output cost of bringing down inflation', *European Economic Review* 18, 85 – 123.

——(1983) 'Real exchange rate overshooting and the output cost of bringing down inflation', in J.A.Frenkel (ed.) *Exchange Rates and International Economics,* Chicago IL: University of Chicago Press.

Clark, K.B. and Summers, L.H. (1982) 'Labour force participation: timing and persistence', *Review of Economic Studies,* 49, 825 – 44.

Cross, R. (1987) 'Hysteresis and instability in the natural rate of unemployment', *Scandinavian Journal of Economics,* 89, 71 – 89.

——(1993) 'On the foundations of hysteresis in economic systems', *Economics and Philosophy,* 9, 53 – 74.

——(1994) 'The macroeconomic consequences of discontinuous adjustment: selective memory of non-dominated extrema', *Scottish Journal of Political Economy,* 41, 212 – 21.

——(1995a) 'Introduction', in R.Cross (ed.) (1995c) *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis,* Cambridge: Cambridge University Press.

——(1995b) 'Is the natural rate of unemployment consistent with hysteresis?' in R. Cross (ed.) (1995c) *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis,* Cambridge: Cambridge University Press.

Cross, R. (ed.) (1995c) *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis,* Cambridge: Cambridge University Press.

Dixit, A. (1980) 'A solution technique for rational expectations models with applications to exchange rate and interest rate determination', mimeo, University of Warwick.

Dornbusch, R. (1976) 'Expectations and exchange rate dynamics', *Journal of Political Economy,* 84, 161 – 76.

Driver, C. (1986) 'The scrapping behaviour of concentrated and non-concentrated industries in the UK', *Applied Economics,* 18, 249 – 63.

Driver, C. and Moreton, D. (1991) 'The influence of uncertainty on UK manufacturing investment', *Economic Journal,* 101, 1452 – 9.

Franz, W. (1987) 'Hysteresis, persistence, and the NAIRU: an empirical analysis for the Federal Republic of Germany', in R.Layard and L.Calmfors (eds) *The Fight against Unemployment,* Cambridge MA: MIT Press.

Friedman, M. (1968) 'The role of monetary policy', *American Economic Review,* 58 (1):1 – 17.

Gocke, M. (1994) 'An approximation of the hysteresis loop by linear partial functions', *Jahrbucher fur Nationalokonomie und Statistik,* 213 (5):572 – 96.

Hargreaves-Heap, S.P. (1980) 'Choosing the wrong "natural" rate: accelerating inflation or decelerating employment and growth', *Economic Journal,* 90, 611 – 20.

Ingham, A., Ulph, A. and Tocker, M. (1988) 'A vintage model of scrapping and investment', *Recherches Economique De Louvain,* 58 (2):169 – 89.

Jenkinson, T.J. (1987) 'The natural rate of unemployment: does it exist?', *Oxford Review of Economic Policy,* 3 (3):20 – 6.

——(1988) 'The NAIRU: statistical fact or theoretical straitjacket?', in R.Cross (ed.) *Unemployment Hysteresis and the Natural Rate Hypothesis,* Oxford: Black well.

Krasnosel'skii, M.A. and Pokrovskii, A.V. (1989) *Systems with Hysteresis,* Berlin: Springer-Verlag.

Layard, R. and Bean, C. (1989) 'Why does unemployment persist?', *Scandinavian Journal of Economics,* 91, 371 – 96.

Layard, R. and Nickell, S. (1986) 'Unemployment in Britain', *Economica,* 53, S121 – S169.

——(1987) 'The labour market', in R.Dornbusch and R.Layard (eds) *The Performance of the British Economy,* Oxford: Oxford University Press.

Layard, R., Nickell, S. and Jackman, R. (1991) *Unemployment: Macroeconomic Performance and the Labour Market,* Oxford: Oxford University Press.

Lindbeck, A. and Snower, D.J. (1986) 'Wage setting, unemployment, and insider-outsider relations', *American Economic Review,* Papers and Proceedings, 235 – 9.

——(1987) 'Union activity, unemployment persistence and wage-employment ratchets', *European Economic Review,* 31, 155 – 67.

Mayes, D. and Young, G. (1994) 'Improving the estimates of the UK capital stock', *National Institute Economic Review,* 147, 84 – 96.

Mayergoyz, I.D. (1991) *Mathematical Models of Hysteresis,* New York: Springer-Verlag.

Nickell, S. (1987) 'Why is wage inflation in Britain so high?', *Oxford Bulletin of Economics and Statistics,* 49, 103 – 28.

O'Shaughnessy, T.J. (1993) 'Hysteresis effects in a model of the Phillips curve', Oxford University Institute of Economics and Statistics, *Applied Economics Discussion Paper Series,* no. 153.

——(1994) 'Hysteresis in an open economy model', Oxford University Institute of Economics and Statistics, *Applied Economics Discussion Paper Series,* no. 156.

Phelps, E.S. (1967) 'Phillips curves, expectations of inflation and optimal unemployment over time', *Economica,* 34, 254 – 81.

Phelps, E.S. *et al.* (1970) *Microeconomic Foundations of Employment and Inflation Theory,* London: Macmillan.

Price, S. (1995) 'Aggregate uncertainty, capacity utilization and manufacturing investment, *Applied Economics,* 27, 147 – 54.

Price, S., Meen, G. and Dhar, S. (1990) 'Employment, capital scrapping and capacity utilization: some results for UK manufacturing', *Discussion Paper* no. 353, University of Essex.

Rowthorn, R. (1995) 'Capital formation and unemployment', *Oxford Review of Economic Policy,* 11, 26 – 39.

Setterfield, M.A., Gordon, D.V. and Osberg, L. (1992) 'Searching for a will o' the wisp: an empirical study of the NAIRU in Canada', *European Economic Review,* 36, 119 – 36.

Soskice, D. and Carlin, W. (1989) 'Medium-run Keynesianism: hysteresis and capital scrapping', in P.Davidson and J.Kregel (eds) *Macroeconomic Problems and Policies*

of Income Distribution: Functional, Personal, International, Aldershot: Edward Elgar.

21
HOW MANUFACTURING CAN HELP YOUNG PEOPLE TO GET HIGH-WAGE JOBS

Roy Green

> [W]hen the momentous events of the late 1980s occurred in Eastern Europe, I was sickened by the ridiculous euphoria, the nauseating complacency and self-satisfaction which emerged in conservative quarters in Western societies. For while I had never been an admirer of the authoritarian, cruel, inefficient and often corrupt regimes that were toppled, by the same token, it could not be said that the performance of the economies of the democratic capitalist West over the last twenty years or so was (or is) anything to write home about either. To have destroyed full employment as a goal, let alone the norm it had become; to have greatly increased the inequality of the distribution of income and of property; to have created an underclass and destroyed the dignity, self-respect and hope of large numbers of citizens; to have substituted ridiculous rewards for paper shuffling for just rewards for making real and useful things; none of these are achievements of which any society could be proud.
>
> (Geoff Harcourt 1992)

INTRODUCTION

Geoff Harcourt, a committed, lifelong social democrat, has always placed himself in opposition to the twin monoliths of free market capitalism and state-dominated socialism. His search for a 'middle way' has focused, in particular, on full employment in a democratic society, and the problem of serious and intractable youth unemployment in the 1980s and 1990s has affected him deeply. This paper examines the problem from the perspective not just of demand management and public sector job creation, but also the role of 'supply side' policy. Essentially, the global shortage of jobs for young people is a shortage of high-wage, high-skill jobs. While in many countries,including Australia, the evidence suggests that employment opportunities are expanding, and will continue to expand, in high-wage, high-skill occupations and sectors, employment growth for young people over the last decade, especially for young

women, has occurred primarily in low-wage occupations. This is especially so where effective job security and minimum wage regulation is lacking.

What can be done to promote high-wage employment for young people? The argument of the paper is that a major expansion of high-wage jobs will depend upon three interrelated elements. The first is a competitive, high-productivity manufacturing sector to create jobs both directly in manufacturing enterprises and indirectly in the rest of the economy through its impact on the balance of payments. The second is a comprehensive training and education system to prepare young people for new technologies and skills associated with high-wage jobs. The third is an industrial relations system which provides for effective minimum wage and equal pay regulation to encourage competition based on quality, design and innovation rather than labour costs, and also to support pay comparability in non-trade exposed sectors. The paper begins by examining the characteristics of fast growing manufacturing economies and the problems and challenges facing Australia. It then proposes a medium- to longer-term 'supply side' policy framework for the expansion of high-wage employment, while maintaining a role for short-term, public sector job creation programmes.

GLOBAL MANUFACTURING IN PERSPECTIVE

What has been happening to global manufacturing? Evidence compiled recently by the Bureau of Industry Economics (BIE) suggests that despite the strong growth of manufacturing output in Australia and the G7 countries between 1970 and 1990, the share of manufacturing in GDP has fallen sharply, reflecting in large degree the accelerated growth in services. By contrast, the GDP share of manufacturing in the Dynamic Asian Economies (DAEs) has risen over the same period. This has occurred primarily at the expense of labour-intensive, low-productivity manufacturing in the G7 countries rather than manufacturing as a whole, which has experienced continuing expansion of output and employment. The G7 countries with the highest share of manufacturing in 1990 were Germany (30.5 per cent) and Japan (29.1 per cent), followed closely by Singapore (29.6 per cent) and Korea (28.9 per cent) in the DAE group. Australia had one of the lowest shares (15.0 per cent), which reflects our continuing dependence upon primary commodity production, and this has been reinforced, as we shall see, by changes in the composition of manufacturing output. The question to be answered is, does this matter? Or putting it another way, what is so special about manufacturing?

The BIE data also indicates that G7 countries with a high share of manufacturing in GDP, and DAE countries with a rapidly growing manufacturing share, recorded the fastest growth of per capita GDP, though the experience of employment growth is more diverse (BIE 1995). Among the G7 countries, the US, which has a relatively low share of manufacturing, has been experiencing the fastest employment growth over the past decade in that group.

However, it also suffers from the greatest inequalities in the labour market (Freeman 1994). By contrast, the European G7 countries, with a higher share of manufacturing, have largely ensured that employment growth does not take place at the expense of equality. The main exception is the UK, which shows signs of achieving the worst of both worlds, having undergone the greatest proportionate fall of any G7 country in manufacturing share of GDP. It not only suffers from the persistently high unemployment of many European countries, but it also shares the US trend to inequality, with the emergence of a two-tier labour market comprising a core of secure, well paid workers and a growing periphery of insecure, part-time, mainly female and young workers.[1] There are other institutional factors that influence these comparisons, such as the role of trade unions and minimum wages, but the long-term economic impact of a presence in manufacturing cannot be ignored. The reasons for this may be briefly summarised.

HIGH-WAGE, HIGH-SKILL JOBS

The significance of manufacturing for growth, employment and real wages is widely acknowledged in the academic and policy literature (Dertouzos *et al.* 1989; Best 1990; Porter 1990; Grossman and Helpman 1991; Sheehan *et al.* 1994). Despite rapid expansion of the services sector in the advanced economies, which substantially accounts for the fall in the share of manufacturing in GDP, elaborately transformed manufactures (ETMs) are still the largest and fastest-growing component of world trade (Green and Genoff 1993). Whereas primary commodities have experienced declining terms of trade over recent decades, hence limiting the benefits of comparative advantage, ETMs have the unique potential to secure what has been depicted as 'competitive advantage' through knowledge-based value adding (Porter 1991). This is the basis for high-wage, high-skill job creation in many successful economies, as well as ecologically sustainable growth, with opportunities for the diffusion of new technologies and skills across industry sectors (Green 1995). Indeed, the evidence suggests that the job displacement effects of new technology, advanced production techniques and work reorganisation tend to be more than offset in the long term by the additional employment created in the manufacturing sector itself, due to the expansion of productive capacity and associated services-based activities. The conclusion of the recent OECD *Job Study* was as follows:

> Technology both eliminates jobs and creates jobs. Generally, it destroys lower wage, lower productivity jobs, while it creates jobs that are more productive, high-skill and better paid. Historically, the income-generating effects of new technologies have proved more powerful than the labour-displacing effects: technological progress has been accompanied not only by higher output and productivity, but also by higher overall employment.

(OECD 1994:33)

Further, it may be argued that the predominant feature of 'post-industrial' society is not that services have replaced manufacturing in the composition of total employment, but rather that the distinction between them has become blurred as manufacturing moves 'upmarket' and contracts out non-core services. Some would claim that this indicates the declining significance of manufacturing in the economy, and draw the parallel with agriculture at the end of the last century, but in reality it reflects fundamental changes in the nature of global manufacturing. While there has been considerable debate about the scope of these changes, it is widely accepted that they encompass a shift, though by no means a universal one, from low-cost, high-volume production units to smaller, more specialised and interdependent units which embody 'flexible manufacturing' techniques (Best 1990; Mathews 1994).[2] These tend to flourish in the 'Marshallian industrial districts' which are characterised by networking and other forms of inter-firm cooperation (Bianchi 1993). In addition, many manufacturing organisations now see themselves as 'a professional service industry, customising [their] offerings to the preferences of special market segments' (Jaikumar 1986:86), including meeting standards of energy efficiency and recyclability. Service organisations have begun to deliver on an increasing scale what may previously have been regarded as manufacturing functions, such as design, maintenance, strategic planning, change programmes and marketing.

As a result, far from declining into irrelevance, manufacturing has taken on an even greater significance for the future of post-industrial economies through its ability to generate major improvements in productivity and competitiveness. This in turn points to a further source of employment growth, which operates through the impact of exports and import substitution on the trade account of the balance of payments (McCombie and Thirlwall 1994). By increasing the 'speed limit' on national GDP growth, manufacturing and manufacturing-related services also contribute indirectly to job creation elsewhere in the economy, including public sector infrastructure and community services. The rate of job creation may then be increased by expansionary demand management policies without triggering a deterioration in the current account deficit. This creates a 'virtuous circle' of growth in output, investment and jobs, which may persist in the longer termeven if employment in manufacturing temporary contracts as a result of firms producing more output with less workers. On the other hand, however, where such a contraction in employment is related not to increased productivity but to the stagnation of manufacturing output and exports, the economy will not benefit from these favourable balance of payments effects. This will result in higher unemployment than would otherwise be the case, and it may also contribute to a 'vicious circle' of high current account deficits, high real interest rates and low levels of investment.[3]

VIRTUOUS CIRCLE OF GROWTH

The main characteristics of the virtuous circle of growth just referred to have recently been identified by the International Monetary Fund (IMF) in a survey of 126 developing countries (IMF 1995). Essentially, the survey compared the performance of the forty-two fastest growing economies with the forty-two slowest growing over the period 1984 – 93, with some cross-references to the 1971 – 83 period. The characteristics of the fastest growing economies included high levels of investment, a rapid rate of productivity growth, rising terms of trade, higher export growth and lower external debt. It is noteworthy that they were also associated with the rapid growth in the manufacturing share of GDP found in the DAE countries. While it cannot be claimed that these characteristics automatically translate into high-wage, high-skill jobs, they do seem to establish the preconditions for expansion in these jobs, just as a declining manufacturing share, if combined with a low rate of capital accumulation and productivity growth, increases the likelihood of a relative contraction of such employment. However, the IMF research does not penetrate very deeply into the causal linkages that give rise to these characteristics, and, to the extent that it makes any attempt to do so, it seeks the comfort of market orthodoxy, highlighting the role of free trade, fiscal restraint and labour market deregulation.

A different analysis of the causal linkages has recently been provided by Rodrik (1995), who has challenged the free market account of the 'growth miracle' in the DAEs with a detailed investigation of the important examples of South Korea and Taiwan. By contrast with the IMF and World Bank, Rodrik argues that exports and 'outward orientation' were not the main drivers of growth, since they represented only a small proportion of national output and had few spillover effects. Nor could it be maintained that export growth was driven by competitive exchange rates, given that real exchange rates in both countries had not changed markedly over thirty years. Rodrik's conclusion from the data is that high levels of investment delivered the export growth, rather than the reverse. He states:

> The proposition that Korea's and Taiwan's economic performance can be ascribed to export orientation faces serious difficulties. The switch towards export-oriented policies cannot account for the sustained export boom since the mid-1960s, and even less for the equally impressive and sustained investment boom. Export growth itself can explain only a limited part of the early growth in output. The increasing share of exports in GDP is quite consistent with a story of investment- rather than export-led growth. These problems should lead us to search for more direct explanations for the apparent increase in the profitability of private investment around the mid-1960s in both countries.
>
> (Rodrik 1985:74 – 5)

Rodrik emphasises the role of intervention in generating the required investment levels, and in so doing rejects the free market view that any government attempt to raise investment must be self-defeating due to its effect on the rate of return. In this view, when the stock of physical capital is in equilibrium, producing a desired rate of return for entrepreneurs, given the cost of financial capital and entrepreneurs' assessment of risk, a government-induced rise in the capital stock will automatically lower the rate of return and impede further investment growth. However, Rodrik's argument lends further weight to the post-Keynesian and Institutional tradition in economic thought, which suggests that this view is flawed both theoretically and when tested against the empirical evidence.

COORDINATION FAILURES

The problem for market orthodoxy is that it assumes that investment markets are perfectly coordinated, bringing the cost of financial capital and the returns from physical investment automatically into an optimal equilibrium position for the economy as a whole. According to Rodrik, this position will not be realised if there are 'coordination failures', where the returns from investment are so low or the returns demanded by finance capital so high that investment remains stuck at low levels. The argument is a familiar one for post-Keynesians like Harcourt, who points out that

> For the patterns of interest rates and prices of financial assets, and the exchange rates which keep overseas holders in particular happy, may well not be the levels which are consistent with the rate of domestic investment spending—both public and private—that may be needed to bring about the desired restructuring and provide theproper level of activity and employment overall, and with which is associated a suitable supply of exports and demand for imports.
>
> (Harcourt 1992:8; see also Cooper and John 1995)

It was precisely such coordination failures that affected South Korea and Taiwan in the 1950s. As both Rodrik and others such as Amsden (1989) have pointed out, these were countries which had highly skilled and educated labour forces, but which lacked an adequate level of manufacturing investment. If all firms could have been persuaded to lift investment, there would have been mutual benefits, but any individual firm acting alone would find the process unrewarding. In addition, investment returns for individual firms were depressed by the cost and difficulty of finding reliable sub-contractors in an only partially industrialised economy, and by the prospect that, having built an expensive production run, demand would be insufficient. Consequently, the economies were trapped in low-output equilibria with poor returns to new investment. The main impetus for growth was a series of interventionist industry policy

measures, which were pursued in the context of relatively open economies, with the price mechanism governing resource allocation, and relatively egalitarian, well educated societies.

The measures included nationalisation of the banking system by the Koreans and the introduction of the Statute for Investment Promotion in Taiwan, which played a part in lowering the cost of capital and lengthening investment time horizons to deal with coordination problems in the financial markets. In Taiwan, a government agency organised the dissemination of foreign technology and best practice, constructing chains of sub-contractors in sectors and industries where Taiwan had no representation. In Korea, the government facilitated the emergence of the *chaebol* conglomerates, in effect allowing each group to construct its own supply chain. There was also a key role here for public investment in creating new industries which the market itself would not have produced, and whose development also generated demand for allied firms and contractors. Again, in Taiwan, glass, cement and plastics factories were established by the state, and then handed over to private entrepreneurs, while the Korean government established POSCO, which is now the world's most efficient steel producer. While these measures comprised an authoritarian element, they demonstrated the crucial role not only of manufacturing in the pursuit of growth and employment but also of alternatives to the market model (see also Weiss and Mathews 1994), including those which encompass more democratic approaches to industry policy.

PROBLEMS AND CHALLENGES IN AUSTRALIA

Historically, Australian manufacturing industry has focused primarily on import substitution and has benefited from high levels of tariff protection. There was little attempt to penetrate global markets for manufacturing, let alone ETMs. However, the tariff reductions of the 1980s and 1990s have opened up Australian domestic markets and provided an incentive, though not in themselves a policy mechanism, for manufacturers to become globally competitive. The result, as we have seen, has been a continuing overall decline in the manufacturing share of GDP, but a remarkable expansion of ETM exports, which grew very slowly between 1979 and 1985 and then tripled between 1985 and 1993. While the net trade deficit in ETMs also increased from $20.8 billion in 1985 to $29.7 billion in 1990, this represents a major decrease in the average annual growth in the deficit and a fall in the ETM deficit/GDP ratio from 9.0 per cent in 1985 to 7.9 per cent in 1990 (Sheehan *et al.* 1994:17 – 19). Even if the deficit continues to rise in current prices, projections for the Centre for Strategic Economic Studies (CSES) suggest that the decline in the ETM deficit/GDP ratio will also be maintained. In addition, the ETM export/import ratio, which fell from about 23 per cent in 1975 to 15 per cent in 1986, had rebounded to over 30 per cent by 1993 and is projected on present trends to reach 50 per cent by 2001. This is still low

by world standards, but the scale of the achievement for a traditional commodity exporter such as Australia is noted in the CSES report:

> If the conditions necessary for the continued improvement in ETM trading performance—continued improvement in the cost base of the economy, further change in business attitudes and more aggressive targeting of policies to support export growth in individual industries—are put in place and the world economy relevant to Australia remains strong, it is possible to envisage an ETM sector in which exports amount to more than one-half of imports. This in turn would make possible such rapid rates of growth as are necessary to lead to the return of full employment, and more generally would generate new areas of competitive economic activity to complement the traditional primary industries and to underpin a viable and prosperous Australian economy in the 21st century.
>
> (Sheehan *et al.* 1994:42)

The CSES report presents the finding that the most rapid growth of ETM exports has been recorded in the six ASIC categories targeted by the Labor Government's sector-based industry policy—pharmaceuticals, computing equipment, telecommunications equipment, road vehicles, other transport and clothing. These are also categories where oligopolistic market structures 'frequently prevail' and where technological and other non-price and qualityfactors are critical to competitive advantage. Of the six categories, four are associated with high-technology products and two are undergoing a major structural adjustment as a response to phased in tariff reductions. The high-tech categories are of particular significance because first, they represent the fastest growing ETM category in world trade, second, they are associated with the most knowledge-intensive products, and third, Australian exports of these products have been increasing at a rate of about 25 per cent a year in both current and constant prices since 1985. Imports have also been growing rapidly, but the current price export/import ratio for these products has risen from 13.5 per cent in 1985 to 32.4 per cent in 1993.[4]

INVESTMENT IN NEW TECHNOLOGIES

The other side of the story, however, is that the improvement in Australia's ETM export performance comes from a low base and still accounts for only a fifth of total merchandise exports. In addition, as Harcourt reminds us, 'we put two separate bits of lead in our saddle bags…by deciding to have a freely floating currency and, following the Campbell Report, a deregulated financial system'. He goes on to argue, with some justification, that

> they made our resulting problems much harder to tackle. Australia needed and needs a very considerable amount of restructuring of its industries. Its traditional export products have not only faced secularly declining demands over the last decade or so, but, for much of the period, demand has been pushed below trend by shorter term cyclical demand deficiencies. This is one of the reasons for the horrendous state of our current account.
>
> (Harcourt 1992:7)

It has also contributed to the roller-coaster ride experienced by the Australian dollar over recent years, with a resulting net depreciation against our major trading partners as well as relatively high real interest rates. While depreciation has made our exports more competitive, it has increased the level of foreign debt which is denominated in other currencies; and while high interest rates attract global capital, they discourage domestic investment.

A national effort will be required to lift Australia's investment performance in manufacturing, which must build upon the success already achieved in a limited range of niche ETM markets and identify new areas of export growth. Yet the rate of structural change in manufacturing as a whole still compares unfavourably with the rate in most G7 and DAE countries, and the evidence suggests that the direction of this change is towards the resource-based, low value-added 'downstream' industries, such as basicmetal products, and away from more complex, knowledge-intensive industries such as fabricated metal. In the G7 and DAE countries, by contrast, fabricated metal products contribute the dominant share of value added and are also growing the most rapidly. This is reflected in the shift in demand away from unskilled jobs, as measured by relative employment and unemployment rates,[5] especially in countries such as Germany and Sweden where real wages growth for low-paid workers has been maintained through industry bargaining and minimum wage regulation. These are countries which also have a long-term commitment to training and vocational education, and are therefore better placed than most to take advantage of new technologies and skills. The implications for unemployment have been drawn by the OECD *Job Study:*

> The appearance of widespread unemployment in Europe, Canada and Australia on the one hand, and of many poor quality jobs as well as unemployment in the United States on the other, have…both stemmed from the same root cause: the failure to adapt satisfactorily to change. Management skills, education and training attainments have failed to keep pace with the requirements of a more technologically advanced economy…. The basic policy message of this report is unambiguous: high unemployment should be addressed not by seeking to slow the pace of change, but rather by restoring economies' and societies' capacity to adapt to it. But this must be undertaken in ways which do not abandon the social

> objectives of OECD societies. Rather, social objectives must be met in new, more carefully designed ways.
>
> (OECD 1994:30)

In its recommendations for tackling unemployment, the OECD covers a number of policy areas, but places particular emphasis on 'adapting education and training systems to a more technology-driven world' (OECD 1994: 34). This is an important consideration, and it is being addressed in various ways by state and federal governments in Australia, but if Rodrik's account of the experience of Korea and Taiwan established anything it was that the primary impetus for exports and employment growth was supplied by public and private sector investment. The correlation between the growth in employment and domestic capital expenditure is also apparent in the Australian and OECD data, though business cycle fluctuations as reflected in this data tend to be much greater for Australia than for the OECD as a whole (Sicklen 1995). Consequently, the decline in the rate of capital formation in Australia must be seen as a key structural factor in the rise of unemployment, along with the cyclical demand factors that precipitated the last three recessions. However, there is no concerted strategy to increase the rate of capital formation apart from the use of wages policy to lowerdomestic cost levels and shift resources from consumption to investment. While this policy has successfully boosted the profit share in national income, it contained no mechanism to direct the resources released by wage restraint into productive investment. Instead, left to the vagaries of the market, those resources were dissipated in the speculative asset boom of the 1980s.[6]

SUPPLY SIDE POLICY FRAMEWORK

The emphasis of the recent debate about unemployment and the jobs outlook for young people has tended to focus on demand side factors, such as public sector employment creation and wage subsidy schemes (Green *et al.* 1992; Langmore and Quiggin 1994). These were identified as a necessary short-term response to the severe impact of a tight monetary policy, which was imposed to rein in high current account deficits and an unsustainable level of external debt, and they also had an independent social and environmental justification. However, the argument of this paper is that the active implementation of demand side policy is also dependent on a coherent supply side framework, whose rationale must be the development of a competitive, knowledge-intensive manufacturing sector. Provided that supply side policy is targeted to lifting investment in productive capacity, it may not only be an important source of jobs itself, but will contribute to overcoming the balance of payments constraint on the growth of output and employment in the economy as a whole. The disappointing aspect of the supply side debate is that while considerable attention has been devoted to microeconomic

reform, labour market flexibility and the national training reform agenda, there is relatively little of substance in current approaches to industry and regional policy, which, despite many reports, have not advanced much beyond the limited industry plans of the mid-1980s.

What form should industry policy take in the 1990s and beyond? While the sectoral dimension of industry policy has been steadily marginalised in favour of more general measures to boost the competitiveness of individual firms, it is important to recognise that the two approaches are not incompatible and may even be capable of acting to reinforce one another. The potential role of sector strategies is highlighted by the spread of networking and customer-supplier links among smaller but more interdependent units of production, and the scope for extending and building upon enterprise bargaining. However, this approach will require new structures to facilitate the development of strategies for each sector by unions and employers, and mechanisms through which the goals and vision of the participants can be translated into reality at the workplace. In this context, the part to be played by factors such as training and education, regional development and EEO policies will be critical, but the essential precondition for high-wage, high-skill jobs remains the level of investment in manufacturing enterprises, especially those exporting ETMs or capable of doing so. Inevitably, low-skill jobs will continue to exist, but the ultimate test of the validity of any policy approach, including this one, will be a shift in the balance of labour market supply to high-skill jobs and a reduction in the high proportion of 'involuntary' part-time employment, which is a form of disguised unemployment.[7]

To sum up, manufacturing can help young people get high-wage jobs through a long-term programme of investment in new technologies and skills, by improving the competitiveness of firms and sectors, by identifying new market opportunities and by vigorously expanding production and exports. This paper has shown that the availability of such jobs is closely related to the state of manufacturing in the economy, and that this role is in turn shaped by the level and quality of investment in high value-added, knowledge-intensive manufacturing products and processes. Many new jobs for young people will also be found in services, but the evidence of this paper demonstrates that a viable, high-skill services sector is similarly dependent upon a dynamic manufacturing base, whether or not those services directly support specific manufacturing activities. It also suggests that the future development of manufacturing industry, particularly ETMs, cannot be left to the market, and that there is scope for planning and intervention, which would encompass, in Harcourt's words, 'information services and back up generally to exporters (and entrepreneurs involved in import replacement) to help them find and then secure niche markets'. He maintains that

> This is an obvious lesson which Australia could learn from those NICs [newly industrialising countries] which gave business people their heads

> but backed them up in the national interest as well. A by-product of being successful in this regard may be a reversal of the trend whereby the 'brightest and the best' were attracted to services and finance sectors by the grossly distorted signals which were given out in the 1980s. Another lesson from the NICs is that we should leave tariff levels where they are, at least in the medium term.
>
> (Harcourt 1993:173)

Finally, high-wage, high-skill jobs in manufacturing, including those for young people, will also be dependent upon the retention of effective minimum wage and equal pay regulation, so that firms and organisations establish their competitive edge through quality rather than wage undercutting. This indicates the continuing importance of an industrial relations system that maintains a balance in its approach to wage determination between the drive for enterprise productivity gains on the one hand and notions of occupational fairness and comparability on the other. It shouldbe mentioned in a paper such as this that the role of fairness in wages policy has also been a long-standing preoccupation of Geoff Harcourt, even when it has meant standing against the prevailing fashion for 'freeing up' the labour market. However, attempts to create a low-wage, low-skill economy through deregulation and the marginalisation of unions threaten not only high-wage jobs but ultimately the low-wage jobs as well. As Harcourt himself insists, 'Australia must not accept unemployment close to 10 per cent as an appropriate "natural" rate of unemployment', which he describes as a 'non-existent concept anyway'. For Harcourt, 'A greater rate of capital accumulation and appropriate macroeconomic policies…can enable a substantial reduction in unemployment over the next five years without a blow-out in the foreign debt or a rapid resurgence of inflation' (Harcourt 1993:174). It is only fitting that the chief beneficiaries of this approach to economic theory and policy will be Australia's young people.

NOTES

1 A third tier has also been identified, comprising an 'underclass' of those excluded from the labour market through unemployment, homelessness, ill-health and poverty (Hutton 1994).

2 There is nothing inherently 'good' or 'bad' about flexible manufacturing. While it may be, and often is, associated with the empowerment of employees through team-based work organisation and the development of permanent stable internal labour markets, it may also result in deskilling, work intensification and the marginalisation of trade unions. The outcome is not inevitable but depends upon factors such as management strategy, union effectiveness and the state of product and labour markets.

3 Harcourt warns that 'if we want markets to work well we must be aware of situations where stocks dominate flows; speculation dominates enterprise or real

economic factors; power is not evenly diffused; prices give out complex signals: and processes are cumulative rather than quickly equilibrating' (1992:5).

4 This confirms OECD research indicating dramatically improved competitiveness for Australian high-tech products, which has the potential for the first time in years to increase the relatively low share of high-tech employment in total manufacturing employment (OECD 1994:19).

5 This trend underlies the argument that job growth in Australia is characterised not so much by a 'disappearing middle' as by a 'vanishing bottom' (Belchamber 1995;cf. Gregory 1993).

6 Harcourt refers to the Accord between the Labor Government and trade union movement as a 'splendid example of conscious and co-operative planning which drew on tried and tested institutions created by our past history and by an appeal to Cupertino from the main classes in society'. This approach

> served to get inflationary pressures very considerably reduced and allowed employment to grow and unemployment to fall from the unacceptable levels of the early 1980s…. The Accord also helped to create the potential surplus for the business classes and the Governments combined to do their part with private investment and the provision of
>
> public infrastructure. That, by and large, this did not occur is not the fault of the Australian wage-earners.
>
> (Harcourt 1992:7 – 8)

7 Australia has one of the highest proportions of involuntary part-time workers and discouraged workers in the developed world (OECD 1995). Indeed, the fragility of the current recovery in Australia is indicated by a comparison carried out by Access Economics with the recovery from the 1980s recession, in terms not of jobs growth but rather in total hours worked. The economy generated twice the hours of work in the earlier period than in the period following the 1990s recession, though the productivity performance as measured by output per hour worked has been significantly stronger in the current recovery.

REFERENCES

Amsden, A. (1989) *Asia's Next Giant: South Korea and Late Industrialisation,* New York: Oxford University Press.

Best, M. (1990) *The New Competition: Institutions of Industrial Restructuring,* Oxford: Polity Press.

Belchamber, G. (1995) 'Disappearing middle or vanishing bottom? Job growth in Australia under the Accord', *ACTU Research Paper,* February.

Bianchi, P. (1993) 'The promotion of small-firm clusters and industrial districts: European policy perspectives', *Journal of Industry Studies,* 1/1.

BIE (Bureau of Industry Economics) (1995) *Australian Industry Trends,* issue no. 22, AGPS, Canberra.

Cooper, R. and John, A. (1995) in G.Mankiw and D.Romer (eds) *New Keynesian Economics,* Cambridge MA: MIT Press.

Dertouzos, M.L., Lester R.K., Solow, R.M., MIT Commission on Industrial Productivity (1989) *Made in America: Regaining the Productive Edge,* Cambridge MA: MIT Press.

Freeman, R. (1994) 'Jobs in the USA, *New Economy,* 1/3.

Green, R. (1995) *The Role and Significance of Metal and Engineering Skills in the Australian Economy,* Sydney: National Metal and Engineering Training Board.

Green, R. and Genoff, R. (1993) 'The challenge of industry policy', in *Making the Future Work: Crisis and Change in the South Australian Economy ,* Sydney: Allen & Unwin.

Green, R., Mitchell, W. and Watts, M. (1992) *Economic Policy in Crisis: A Proposal for Jobs and Growth,* Sydney: Evatt Foundation.

Gregory, R. (1993) 'Aspects of Australian and US living standards: the disappointing decades 1970 – 90', *Economic Record,* March.

Grossman, G. and Helpman, E. (1991) *Innovation and Growth in the Global Economy,* Cambridge MA:MIT Press.

Harcourt, G. (1992) 'Markets, madness and a middle way', The Second Donald Home Address, Melbourne: National Centre for Australian Studies, Monash University.

——(1993) 'Macroeconomic policy for Australia in the 1990s', *The Economic and Labour Relations Review,* 4/2.

Hutton, W. (1994) *The State We're in,* London: Jonathan Cape.

IMF (International Monetary Fund) (1994) *World Economic Outlook,* Washington DC: IMF.

Jaikumar, R. (1986) 'Post-industrial manufacturing', *Harvard Business Review,* November-December.

Langmore, J. and Quiggin, J. (1994) *Work for All: Full Employment in the Nineties,* Melbourne: Melbourne University Press.

Mathews, J. (1994) *Catching the Wave,* Sydney: Allen & Unwin.

McCombie, J. and Thirlwall, A. (1994) *Economic Growth and the Balance of Payment Constraint,* New York: St Martin's Press.

OECD (1994) *Job Study: Facts, Analysis, Strategies,* Paris: OECD.

——(1995) *Employment Outlook,* Paris: OECD.

Porter, M. (1990) *The Competitive Advantage of Nations,* London: Macmillan.

Rodrik, D. (1995) 'Getting interventions right: how South Korea and Taiwan grew rich', *Economic Policy,* April.

Sheehan, P., Pappas, N. and Cheng, E. (1994) *The Rebirth of Australian Industry,* Melbourne: Centre for Strategic Economic Studies, Victoria University.

Sicklen, D. (1995) *A Fresh Look at Industry and Trade Policy: Some Preliminary Suggestions,* Sydney: Australian Economic Analysis.

Weiss, L. and Mathews, J. (1994) 'Innovation alliances in Taiwan: a coordinated approach to developing and diffusing technology', *Journal of Industry Studies,* 1/2.

22
HARCOURT, HICKS AND LOWE
Incompatible bedfellows?

Peter Kriesler

> 'Well, in *our* country,' said Alice, still panting a little, 'you'd generally get to somewhere else—if you ran very fast for a long time, as we've been doing.'
>
> 'A slow sort of country!' said the Queen. 'Now, *here,* you see, it takes all the running *you* can do to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that.'
>
> 'I'd rather not try, please!' said Alice. 'I'm quite content to stay here.'
>
> (The Red Queen's explanation to Alice of the principles of the traverse, in Lewis Carroll, *Through the Looking Glass* [1872])

INTRODUCTION

In a very interesting and important paper titled 'Marshall, Sraffa and Keynes: incompatible bedfellows?' Harcourt explores the nature of centres of gravitation as explanations of long-period equilibrium. In that paper, he provides an excellent taxonomy for the possible meaning of these positions, as well as outlining some of the problems associated with them.

The paper examined the nature and taxonomy of long-period positions, before providing, by building on the foundations laid by Joan Robinson, the basis of an important critique of both the static method and of equilibrium analysis:

> We then are required to scrap the short cut of using the statical method and tell a much more sophisticated story of the initial failure to reach an implied rest state changing the rest state itself.
>
> (Harcourt 1981:216)

> Exactly the same puzzles hound the concept of prices of production, especially when we try to incorporate them as operational concepts in an analysis of modern oligopolistic economies. It is not, as orthodox economists would argue, that the existence of a tendency to equality of

> rates of profit in all activities may be questioned because of oligopolistic structures, barriers to entry and all the other paraphernalia of modern I.O. It is, rather, that the dynamic nature of capitalist development with the embodiment of technical advances through investment expenditures is so rapid in most periods as not to allow sufficient historical time for centres of gravit[ation] of a lasting nature to be formed. There is not the time, as Joan Robinson puts it, for the traders to become familiar, through actual experience, with what is the norm, so that when their bearings are cut loose, they—and the economists, too—literally are all at sea, rudderless, not knowing where they are heading, either back or to. The factors that we need theoretically to take as constant in order to allow the centres of gravit [ation] which they imply to be struck (for example, by the forces making for the formation of normal prices) are changing as fast as, or even faster than, the outcomes that the relationships between them are intended to determine.
>
> (Harcourt 1981:218)

We have here the basis of a fundamental critique not only of the viability of long-period positions, but also of the nature of equilibrium analysis. Harcourt's critique points to the argument that the forces which push the economy to its long-period equilibrium will also change that equilibrium. In other words, the adjustment path will change the structure of the economy, thus influencing the final equilibrium position. This suggests either that the concept of equilibrium has no operational meaning, or that equilibria, when analysed properly, must be path-dependent.

This paper continues the investigation into the nature of equilibrium analysis developed in Harcourt's paper, by exploring the concept of the traverse,[1] which relates to the question of adjustment of the economy to equilibrium, that is, to 'disequilibrium' adjustment.[2]

The traverse is at the same time one of the most important concepts in economic theory, and also one of the most neglected.[3] Modern economic theory is normally concerned with some concept of equilibrium, and with properties of equilibrium systems. Other than lip service in the form of stability analysis, very little is said about the path an economy may take to reach an equilibrium, if it is not already there. In this paper we are concerned with this question in the form of the relation between disequilibrium analysis and the concept of the traverse. According to Hicks, who first coined the term,[4] the 'traverse' traces out 'the path which will be followedwhen the steady state is subjected to some kind of disturbance.' (Hicks 1973: 81). In other words, the traverse describes the dynamic (out of equilibrium) adjustment path in historical time.[5] Some economists have been very vocal about the fruitlessness of studying the equilibrium properties of an economic system without considering the question of whether the economy will actually get there. In other words, they have voiced

their doubts about the comparative static method which dominates modern economics.[6] More pointedly, Joan Robinson (as well, of course, as Harcourt) has often criticised the separation of equilibrium analysis from the analysis of the traverse, as she believed that the actual equilibrium which an economy achieves (if it is capable of achieving one)[7] will be vitally dependent on the path it takes.[8]

A TAXONOMY

It is important to distinguish between the concept of equilibrium and the use of equilibrium analysis. Equilibrium, as an organising concept, plays an important role in most economic theory, with no implications about an economy's ability to achieve that position.[9] This differs from 'equilibrium theory', which utilises the comparative static method to compare equilibrium positions. Basically, we can distinguish three views as to the role of the traverse and of disequilibrium analysis in economic analysis. The most widely held view is that which analyses the economy in terms of static equilibrium positions utilising the comparative static method, but accepts that the economy may not always be in equilibrium. Within this view, we can distinguish three substreams. In the first of these, the analysis of out-of-equilibrium positions is relegated to minor importance, if discussed at all. By assuming that the equilibrium is stable, the issue of the actual path the economy may take is ignored. The second view, which is by no means incompatible, is that shared by the Classical economists, in their value theory, as well as by some neoclassical economists.[10] According to this view, if an economy is out of equilibrium then it will be attracted to a given equilibrium position, which remains constant, regardless of the path it takes to approach it. Many neoclassical economists who simply consider the sufficiency conditions for the stability of a general equilibrium can be seen to be within this stream. Finally, there are those economists who, while they accept that an economy will tend towards equilibrium, argue that the final equilibrium position is path-determined. There are two interpretations of 'path determinacy'. The first occurs where there are multiple equilibria, so that the one which the system actually achieves is dependent on the path it takes when not in equilibrium. This is contrasted with the path dependency associated with hysteresis which:

> is due to the fact that the movement of the system when it is out of equilibrium may change the data on which the static equations which define the equilibrium are based, so that these equations will change and determine a *different* equilibrium and so on and so forth. In other words, the (set of) equilibrium point(s) is *not* independent of the dynamic movement of the system, that is, this set is path-dependent.
>
> (Gandolfo 1987:461)

It is this latter interpretation of path determinacy which is the main concern here, with the later works of John Hicks and the works of Adolph Lowe being of particular relevance. Some neoclassical economists also implicitly share this view by showing that there may not be a unique equilibrium, in which case the actual equilibrium achieved will be path-determined. In addition, recent work on hysteresis,[11] particularly in the labour market[12] can be considered within this view as the 'equilibrium' rate is seen to be dependent on the length and severity of any deviation from it, i.e. it is path-determined.

The second view of the role of the traverse is peculiar to the new Classical macroeconomists. According to this view, there is no role for the traverse as all we need to know is subsumed in the equilibrium state.

The final view is that an economy is never in static equilibrium, nor does it tend towards it. In this case, dynamic analysis without reference to static equilibrium would be of utmost analytical value. This is the view of those economists who eschew the use of *conventional* equilibrium analysis. The best known of these is Kalecki.

It is important to note that some economists hold one of these views with respect to certain variables, but another with respect to others. For example, it may be felt that there is some equilibrium level of prices, but no corresponding equilibrium with respect to output. Some would put Adam Smith in this category, since his analysis of price is usually in terms of long-period centres of gravitation, but his analysis of growth in output is in terms of dynamic increasing returns to scale.[13] In addition, there are many who belong in more than one stream. In particular, many neoclassical as well as Sraffian economists doubt that the necessary stability conditions are fulfilled.[14] Moreover, the work of Lowe can also be considered within this hybrid group, as he considers the likelihood of the traverse converging to a new equilibrium.

This paper looks at these three views and considers the manner in which they treat the question of the traverse. The first group by far dominates the discipline, for epistemic rather than scientific reasons. The second view emerged as a historical reaction to the advances of Keynesianism. Finally, the third view, namely that the economy is always on a traverse, can be associated with the works of Kalecki and Joan Robinson; the contributions ofHicks and Lowe to the analysis of the traverse will be considered in the same section.

EQUILIBRIUM AND THE TRAVERSE

The Classical economists analysed the forces which would bring the economy back to what they referred to as natural positions. Rather than using the concept of equilibrium, they analysed the economy in terms of 'centres of gravitation' (à la Smith). In their analysis, the traverse describes the adjustment process of the economy towards these centres. The forces which brought prices back to their natural positions were normally conceived in terms of a tendency towards a

uniform rate of profits. This tendency, coupled with the free mobility of capital, meant that mobile capital responded to any differential from the uniform rate of profits.[15] That is, if, for example, a particular sector of the economy was earning larger rates of profit than the average, then capital would move into that sector, increasing supply, and reducing both price and profit rates. This adjustment would continue until the profit rate in that sector was brought into line with the average rate of profit. So, for the Classical economists, the traverse was propelled by the tendency towards uniform rates of profit which ensured that market price would gravitate towards natural prices. In their vision of the determinants of the centres of gravitation of prices (that is, natural prices), demand played no direct role. For this reason, the analysis of what happened when there was a divergence between natural and market prices concentrated on 'actual' values rather than, as with most stability analysis, 'conjectural' values. In other words, because demand does not play an important role in the determination of natural prices, trading outside natural (equilibrium) prices is allowed within the stability analysis. This means that the Classical economists were analysing a dynamic system with moving centres of gravitation in their determination of value, with the analysis occurring in historical time. The forces which produced the convergence to natural positions were different to the forces which determined those positions, so that they were not path-determined.[16]

The modern version of the Classical theory, labelled alternatively Sraffian or neo-Ricardian economics, attempts to resurrect this method. However, there are underlying problems which have arisen since the time of the Classical economists related to the changing nature of capitalism, which are not addressed. In particular, the adjustment to the long-period position is either simply subsumed under the term 'tendency towards a uniform rate of profit', or else the adjustment path is shown to converge under some conditions. This last approach involves a fundamental fallacy. It is not enough to show that convergence to long-period positions can occur under some specific conditions. For the long-period approach to be methodically valid asan analytical framework for the analysis of capitalist economies, it must be shown that such convergence will occur under *all* reasonable conditions. In any case, it is difficult to maintain the validity of the analytical separation of the forces determining equilibrium from those pushing the economy to that equilibrium (as the discussion of Harcourt noted above indicates) for modern capitalist economies. This is due to the fact that the investment process, which is supposed to equalise profit rates between sectors, does not merely change the quantities of commodities produced, as it did for the Classical economists. Rather, since in modern capitalist economies investment is embodied in machines, investment is associated with technical progress and increasing returns in a cumulative manner. This, of course, as Young, Kaldor and Myrdal have shown, is extremely destructive of any concept of equilibrium. In any case, it is difficult to argue that forces pushing equalisation of profit rates are the dominant forces in modern monopolistic economies.[17]

The analysis of the early neoclassical economists was comparative static, although there was some lip service paid to disequilibrium-type problems in the form of stability analysis. For Walras, stability was brought about by the relationship between price movements and excess demand. Walras postulated that if any market was not in equilibrium at the current price, then price would move in the same direction as excess demand. Unlike the Classical economists, whose analysis allowed economic activity to continue even when market prices deviated from their natural values, in Walras' analytical framework transactions are ruled out by axiom until equilibrium prices have been determined, due to the role of demand in the determination of those equilibrium prices. No trading was allowed out of equilibrium, since if there were trading, then it is analytically unfeasible to obtain an equilibrium solution. This was due to the change in the value of agents' initial endowment if trading at non-equilibrium prices occurred. If trading outside equilibrium is allowed to occur then the system will exhibit hysteresis. To overcome this problem Walras introduced the infamous 'tâtonnement' process,[18] while Edge worth utilised the notion of 'recontracting'. Their equivalent of a traverse was, in effect, conjectural analysis in the sense that no actual exchanges were transacted until an equilibrium position was established. These models could not trace out the actual adjustment path of the economy, but rather described sufficient conditions for an equilibrium to be stable, but only in the 'conjectural' sense described above. This approach corresponds to Joan Robinson's notion of 'logical time'.[19]

Modern neoclassical economists have studied the stability requirements of a general equilibrium system at great length. The first point to note is that the analysis usually suggests the possibility of multiple equilibria. If this is the case, then the actual equilibrium reached will be dependent on the path the economy takes when it is out of equilibrium, as well as its initial position. However this should not be confused with the hysteresis discussedearlier, as the actual data of the system, in this case, *are not changed* by the movement outside equilibrium. In the case of multiple equilibria, the position of the equilibrium is determined independently of the adjustment path. The path will only determine which equilibrium the economy tends towards. Otherwise, modern stability analysis is usually conducted using the same methodological approach as that of its predecessors. The analysis is about the sufficient conditions for stability, given that no trading occurs until an equilibrium position is reached. Where trading outside equilibrium is permitted, very stringent postulates are required to show that an exchange economy converges to an equilibrium, and the system will exhibit hysteresis.[20] Difficulties of showing convergence within production economies may be highlighted by the work of Hahn, who, after considering many dynamic production models, appears highly sceptical about the possibility of economies, as described by modern general equilibrium theory, being able to reach, much less maintain, a steady state (equilibrium) growth path (see Hahn 1985: Parts III and IV).[21]

It should be noted that there are aspects of modern neoclassical analysis in which the path-determined nature of the equilibrium is explicitly analysed. One such area is that of chaotic dynamics, which has only recently been incorporated into economic theory.[22] There is also the work explicitly associated with the concept of hysteresis, especially in the labour market. The basic idea behind such models is that the 'equilibrium' level of unemployment, as well as the trajectory of employment, will be determined by the path which the economy takes when it is out of equilibrium.[23] In particular, a level of unemployment greater than the equilibrium level will tend to raise the 'natural' rate, while a level of unemployment lower than the equilibrium level will tend to lower it. Various explanations have been proposed to explain this phenomenon, with most focusing on the process of wage determination within the labour market.

Before continuing, it is important to consider the limitations of the comparative static method which underlies much of the analysis discussed so far. The validity of this method rests on a number of assumptions about the underlying dynamics which the economy must exhibit outside equilibrium. First, it must be shown that the economy is stable; that is, if it is displaced from equilibrium then there are forces which ensure that it converges to an equilibrium position. Second, equilibrium can not be path-determined in the sense of hysteresis, for if it is, then in order to determine the new equilibrium, the dynamic adjustment path must be traced out. The result of not accepting at least one of these properties must be a rejection of the comparative static method.

In addition, for that method to generate useful insights, the convergence to equilibrium must be relatively 'rapid':

> If the predictions of comparative statics are to be interesting in a world in which conditions change, convergence to equilibrium must be sufficiently rapid that the system, reacting to a given parameter shift, gets close to the predicted new equilibrium before parameters shift once more. If this is not the case, and, a fortiori, if the system is unstable so that convergence *never* takes place, then what will matter will be the 'transient' behaviour of the system, as it reacts to disequilibrium. Of course, it will be a misnomer to call such behaviour 'transient', for it will never disappear.
>
> (Fisher 1983:3, original emphasis)

Of course, modern neoclassical economists have attempted to grapple with these issues, and Hahn, Negishi and Fisher processes are some of the fruits of these works. However, these all involve an abandonment of the comparative static method.

THE IRRELEVANCE OF THE TRAVERSE

A group of economists sometimes called the 'New Classical Macroeconomists' have popularised a radical new method for analysing economic behaviour which effectively rules out disequilibrium and therefore the traverse. Essentially, the 'New Classical Macroeconomics' (NCM) arose as an answer to the sterile debates of the 1960s between neoclassical Keynesians and monetarists as to the efficacy of monetary policy. Basically, both groups agreed that the economy tended towards a long-run full employment equilibrium, but the 'Keynesians' argued that there were short-run impediments which may mean that in the short term unemployment may persist for some time. The main culprit in the story was expectations. The NCM reply to this was to assume that expectations were determined in the same way as other economic variables, that is, by rational economic agents making optimal use of the resources and information available to them. In effect, what they assume is that economic agents 'are aware of the values of the variables affecting the market where they currently are…and of the true probability distributions governing the future state of this market and the present and future states of all others' (Lucas 1983:158). The result of this assumption, coupled with the postulate of continual market clearing, is in effect to abolish the distinction between the short run and the long run, in that, if agents are aware of the equilibrium values of all variables, then they will always act on that information, and so will act to ensure that 'prices and quantities are taken to always be in equilibrium'. (Lucas 1983: 287, also 179), So, 'New classical economists defy the convention and interpret the equilibrium price as the actual price' (Klamer 1984:15). Even the business cycle is regarded as an equilibrium phenomenon.[24] These aredefined as equilibrium models because, 'in these models, the concepts of excess supply and demand play no observational role and are identified with no observational magnitudes' (Lucas 1983:287).[25]

Clearly, since the analysis assumes that agents are always and everywhere in equilibrium there is, by definition, no possible role for the traverse.

LIFE IS A TRAVERSE

This section will consider the work of Hicks and Lowe on the economic analysis of the traverse, as well as discussing the contributions of Kalecki to dynamic analysis. All these writers, in the works under consideration, eschew comparative static analysis in favour of a dynamic analysis in historical time. In all the models, the sequence in which events occur is important and irreversible. Furthermore, they shun the method of analysing individual agents, which is manifest in the neoclassical models discussed. In their place, they focus either on the structure of production, or in the case of Kalecki, on the class analysis of effective demand.

As was noted above, Hicks was the first economist to use the term 'traverse' in his *Capital and Growth* (1965). Here Hicks studies the various methods

economists have utilised for analysing economic dynamics. After considering the concept of 'growth equilibrium', Hicks turns to the question of the traverse:

> Suppose we have an economy which has in the past been in equilibrium in one set of conditions; is it possible (or how is it possible) for the economy to get into a new equilibrium, which is appropriate to the new conditions? We do not greatly diminish the generality of our study of disequilibrium if we regard it in this way, as a Traverse from one path to another.
>
> (Hicks 1965:184)

Hicks' analysis of the traverse is in terms of a two-sector, fixed coefficients model. The model represents his attempt to utilise the methods of the Classical economists, and is very similar to Sraffa's model,[26] with two important differences. First, because Hicks assumes that capital is immortal, there is no depreciation and therefore he can avoid joint production. Second, Sraffa's model is a static model whereas Hicks' is a steady-state growth model. In part, the return to the method of the Classical economists, whose analysis of growth are reviewed by Hicks in the earlier parts of that book, represented attempts to look at the questions of growth in terms of dynamic equilibria. Hicks explicitly acknowledges that he is attempting to follow the path derived from Harrod and Domar and developed by Joan Robinson, Kaldor, Samuelson and Solow.[27] The model differs from that of Lowe andKalecki in that it considers a two-sector model with one capital good which can freely be moved between capital and consumption goods sectors. Without the complications implied by structural disproportionalities, Hicks concludes that a full-employment path to equilibrium is only possible if the consumption goods sector is more mechanised than the capital goods sector. Even if this condition is fulfilled, a full employment traverse is not guaranteed, but must satisfy a series of technologically determined conditions with respect to the man/machine ratios in the two sectors (Hicks 1965: 187 – 90). Hicks reaches the important conclusion that 'smooth adjustment may not be possible' (*ibid.:* 190).[28] In *Capital and Time* (1973), Hicks moves away from the 'Classical' traverse, and attempts to analyse the traverse within an Austrian framework. In order to get a unique and unambiguous period of production, Hicks has to resort to the uninteresting case of the 'simple profile', for reasons related to the capital controversies (as he makes clear at 41 – 4). Unfortunately, this becomes essentially a one-product world, and is not, therefore, particularly enlightening. This latter attempt of Hicks to analyse the traverse within a neoclassical framework may be contrasted with the efforts of Kalecki and Lowe, and illustrates the difficulty of using the neoclassical approach to meaningfully discuss disequilibrium phenomena.

Lowe,[29] in a return to the concerns of the Classical economics, sees the main problem of economics as the description of the 'path of economic growth'. He believes that growth is not normally of the steady-state equilibrium type, so he

specifically analyses the traverse, and concentrates on the implications of structural change. For this reason, he focuses on the nature of changes in the structure of production and on intersectoral relations, again reminiscent of the Classical economists. To examine this problem, Lowe developed a three-sector model which incorporated not only the concept of historical time but also two important aspects of production rarely dealt with by modern economists. These were the specificity of capital goods,[30] and the importance of reproduction, which is necessary for the incorporation of intersectoral relations. Specificity is dealt with by differentiating two subsectors of the capital goods sector (sector 1). In the first subsector (1a) capital goods are produced which can either reproduce themselves or produce capital goods for the consumption goods sector (sector 2). Although, at this stage, there is no distinction between the capital goods, specificity becomes important when the capital goods produced in this sector are installed so as to produce capital goods for the consumption goods sector. On installation these capital goods lose their generality and, in an irreversible process, become specific to the production of capital goods for the consumption goods sector. These capital goods may be considered a separate branch of production (1b). The capital good output of this sector is installed in the consumption goods sector to produce consumption goods.

Lowe uses this model to examine the nature of the traverse. He begins by assuming that the economy is initially in a stationary state, and considers theimplications for the traverse of changes and restrictions on variables. The model is used to consider the structural changes within the capital goods sector which are necessary to facilitate, for example, changes in technology and changes in the rate of growth of the labour force. In this way the analysis is able to consider changes in the structure of production and their implications for intersectoral relations during the traverse, and so consider the likelihood of a new steady state emerging.

Without going into the detailed mechanics of the model, an important conclusion to emerge is Lowe's demonstration that, although there may very well be a traverse which leads to a new full employment steady state, it is unlikely to be achieved within a decentralised market system. This, in part, results from the market transmitting the 'wrong' signals in terms of the optimal structure of production and intersectoral flows.

For Kalecki, like Lowe, the object of analysis of modern capitalist economies is the dynamic growth path of the economy. However, in contrast to Lowe's work, Kalecki does not use static equilibrium or the steady state, even as reference points. Rather, he explicitly eschews equilibrium analysis, casting doubt on its validity as a method appropriate for analysing growth in modern capitalist economies.[31]

Kalecki's central concern was with understanding the movements of actual economies, in particular, with the analysis of business cycles and growth. Most of his other economic analysis can be interpreted as steps toward this final goal.[32]

Although Kalecki, like Lowe, disaggregated the economy into three sectors, or departments, the disaggregation served different purposes. Kalecki divided the economy into one capital goods sector and two consumption goods sectors, differentiating worker's consumption from that of capitalists. The distinction from Lowe's model reflects Kalecki's different emphasis; rather than being concerned with the structure of production, Kalecki was concerned with problems associated with realisation in the form of effective demand, as he believed that The main problem of a developed capitalist economy is the adequacy of effective demand' (Kalecki 1976: 20).[33]

He did, however, share Lowe's concern with reproduction and with intersectoral relations, but concentrated on flows of commodities and of incomes between sectors. As a result of Kalecki seeing the main determinant of income and growth in mature capitalist economies as being the level of effective demand, he made it one of the central elements of his analysis. Related to this was the role of investment, which he perceived as having a dual aspect. On the one hand investment was part of effective demand so that the higher the level of investment in any period, the greater the level of employment in that period. On the other hand, because investment contributed to the creation of extra capacity, the higher was the level of investment in this period, the larger would be the problem with achievingfull employment in the next. This 'paradox', according to Kalecki, struck at the heart of the capitalist system: The tragedy of investment is that it causes crisis because it is useful. Doubtless many people will consider this theory paradoxical. But it is not the theory which is paradoxical, but its subject—the capitalist economy' (Kalecki 1939:148 – 9).

Although Kalecki concentrated on the role of effective demand in his analysis of capitalist economies, in his work on socialist economies the structure of production, rather than effective demand, was seen as an important constraint on economic activity. Here he came much closer to the traverse analysis of Hicks and Lowe, and in many ways their efforts are complementary to Kalecki's. In the analysis of the multiplier and the effect of a reduction in investment on the level of economic activity, Kalecki differentiated between the implications for capitalist economies and for socialist economies. In a capitalist economy, a reduction in investment causes a reduction in profits which feeds through to a multiplied reduction in income, due to the resultant reduction in consumption. Kalecki contrasts this to the effects of a reduction in investment in a socialist economy, where he argues there is no problem with effective demand:

> The workers released from the production of investment goods would be employed in the consumption goods industries. The increased supply of these goods would be absorbed by means of a reduction in their prices. Since profits of the socialist industries would be equal to investment, prices would have to be reduced to the point where the decline in profits would be equal to the value of the fall in investment.

(Kalecki 1954:62 – 3)

Notwithstanding his contention that a change in investment in socialist economies would lead in the long period to shifts in the structure of production without implications of the multiplier on the level of economic activity, and hence without the problems with effective demand that occurred in capitalist economies, Kalecki accepted that the short-run adjustment process could be problematic. In his analysis of 'the structure of investment in socialist economies' (Kalecki 1963), he acknowledges the possibility of short-run problems in adjustment caused by capacity bottlenecks, in the sense of too much (or too little) capacity in the capital goods sector. It is here that his work touches on issues raised by Hicks and Lowe. Using a two-sector model, and differentiating investment in the capital goods sector from aggregate investment, he shows that changes in the growth rate of the economy will necessitate deviations between the growth rate of investment and that of the economy, during the transition period. However, 'there exists a ceiling to the deviation of the rate of growth of investment from that ofnational income which is determined by the production capacity of the investment sector' (Kalecki 1963: 107).[34]

Kalecki's two-sector model suffers from its inability to sufficiently disaggregate the structure of the investment goods sector. This is of particular importance when the problem is that of differential growth rates between the consumption and investment goods sectors, with investment goods being provided to both sectors. Kalecki acknowledged the problem by allowing for changes in the way equipment is used, especially 'the possibility of turning plant used in the manufacture of consumer durables to production of machinery' (Kalecki 1963: 109); and through favourable changes in the composition of foreign trade. The analysis is reformulated, introducing a coefficient to account for these effects, and this causes the capacity constraint to be relaxed. Halevi (1981) shows that the meaningfulness of this coefficient depends on implicit assumptions about the nature of the capital equipment in the two sectors, and that such implicit assumptions are not necessary in a Lowe three-department model. It is here that Lowe's model can be seen as supplementing Kalecki's discussion of structural problems in the investment goods sector, as well as showing the difficulty of getting rid of excess capacity.

Finally, as well as a sceptical attitude towards the relevance of equilibrium method to the study of modern capitalist economies, Kalecki was also dubious about the validity of the 'long-period' method. In particular, he argued: *'In fact, the long-run trend is but a slowly changing component of a chain of short-period situations; it has no independent identity'* (Kalecki 1971: 165, emphasis added).

This would appear to contradict the Classical notion of 'long-run' centres of gravitation, from which market price may deviate in the short period, but will

gravitate towards. In other words, it is a clear denial of the validity of long-run method, *a fortiori* long-run equilibrium.

This provided a particularly important influence on the works of Joan Robinson. From her earliest criticisms of neoclassical capital theory (Robinson 1953 – 4) she has been very critical of analysis which uses an equilibrium methodology, stressing that such a methodology must, by its very nature, ignore both history and uncertainty about the future. In particular, Robinson has always stressed the inability of comparative static analysis to handle the essence of modern capitalist economies, historical time. In doing so, she has denied the possibility of economies actually getting into equilibrium if they are not already there:

> The neo-classical economist thinks of a position of equilibrium as a position towards which an economy is tending to move as time goes by. But it is impossible for a system to *get into* a position of equilibrium, for the very nature of equilibrium is that the system is alreadyin it, and has been in it for a certain length of past time.
>
> (Robinson 1953 – 4:120, original emphasis)[35]

CONCLUSION

This brief survey of the role of the traverse in economic theory has indicated its importance for almost all economic theory, as signalled by Harcourt's comments noted above. At the same time it has pointed to the sad state of neglect of the traverse in modern theory, especially neoclassical theory. Notwithstanding this, it clearly has an important role, even in equilibrium theory. Unless it can be shown that any disturbance from an equilibrium will lead to an actual (rather than 'conceptual') traverse which converges to a new equilibrium relatively quickly and without hysteresis, the whole edifice of comparative static methodology is fundamentally flawed. There has been much analysis which shows that any deviation from an equilibrium position will lead to cumulative movements away from that equilibrium. Important examples are the concept of cumulative causation made popular by Wicksell and Myrdal, and revived by Young in his work on increasing returns (Young 1928), which in turn inspired Kaldor's work on Verdoorn's Law (Kaldor 1966). Furthermore, many other writers have considered dynamic non-equilibrium models. Goodwin, for example, has made important contributions in this field. In particular, his predator/prey model of a growth cycle and his dynamic non-linear analysis (reprinted in Goodwin 1982) are important contributions to the field. However, the work of these economists remains very much outside the mainstream, and has had little impact.[36]

There is still much to do.

ACKNOWLEDGEMENTS

I would like to thank, but in no way implicate, Neil Hart of the University of Western Sydney, Bill Junor of Macquarie University, Joseph Halevi of the University of Sydney, Geoff Harcourt of the University of Cambridge, Victoria Chick of University College London, Peter Reynolds of Staffordshire University, Greg Mahony of the University of the West of England, and Bruce McFarlane for their helpful comments.

NOTES

1 An example of Harcourt's discussion of the role of the traverse in non-neoclassical theory is in Harcourt 1979:171 – 4.
2 'It is only by a study of "disequilibrium" adjustment that we will get adequate answers to the [fundamental questions of economics]' (Harcourt 1976:136).
3 For example, in the *New Palgrave* dictionary (Eatwell *et al.* 1987), there is no entry on Traverse.
4 As far as this author is aware; see also Lowe (1976:10n).
5 For the distinction between 'historical' and 'logical' time see Robinson (1953, 1974). The importance of historical time is taken to be its unidirection, so that time can only move forward. This implies that the link between time periods is given by the stock of capital inherited from the past, and the expectations determined by it.
6 This is also true of some neoclassical economists. See, for example, Fisher (1983).
7 She was always dubious of the possibility of an economy *ever* achieving equilibrium. See below, final section.
8 See, for example, Robinson (1953, 1974).
9 For an excellent discussion of the way in which the concept of 'equilibrium' is dependent on the theoretical approach being used, see Dow (1985: ch. 5).
10 It should be noted that we do not consider all neoclassical theory, as such a task is beyond the scope of this paper. Rather the focus is on the underlying 'core' of neoclassical theory, the comparative static models of intertemporal general equilibrium.
11 The term 'hysteresis' is defined as 'referring to situations where equilibrium is path-dependent' (Blanchard and Summers 1987:289).
12 There has also been some work which extends the analysis of hysteresis to the study of international trade and to the study of firms' fixed costs. See Baldwin (1988, 1989).
13 Lowe (1975) describes Smith's analysis as a dynamic growth process with 'dynamic feedbacks' and a 'spiralling path of growth'.
14 A good example of this within neoclassical theory is the textbook analysis of 'cobweb' cycles in agriculture, where the adjustment process may either converge to equilibrium or explode.
15 It should be noted that, despite superficial similarities, this process is quite different from the neoclassical concept of adjustment in a perfectly competitive market. See Eatwell (1982) and Harris (1988).
16 Cf. Harcourt (1981).

17 These arguments are spelled out in greater detail in Halevi and Kriesler (1991).
18 For a detailed explanation of two different interpretations of the tâtonnement process, see Hahn (1985:192).
19 See Robinson (1974).
20 See, for example, Arrow and Hahn (1971: ch. 13) and Fisher (1983).
21 In addition we can mention modern monetary analysis of the exchange rate, where the analysis is an outgrowth of control theory, which considers the adjustment of the exchange rate and prices to monetary disturbances. The analysis allows for initial overshooting, due to the role of expectations, before eventual adjustment to the new equilibrium, which is independent of the adjustment path (see Dornbusch 1976).
22 For a survey of recent economic developments of chaotic theory see Frank and Stengos (1988).
23 The notion of the trajectory and path being dependent on the direction of movement is a result in catastrophe theory known in economics (at least) since Zeeman's 1977 presentation (Zeeman 1977).
24 See Lucas (1983:179 – 240, 271 – 97); and Lucas in Klamer (1984:40 – 2).
25 See also Lucas in Klamer (1984:38). For the definition of equilibrium in NCM see Sargent in Klamer (1984:68 – 9).
26 See Hicks (1985:132n).
27 Hicks (1965:vi). The importance of Robinson (1965) is explicitly noted.
28 For the inappropriateness of prices as a guide to decision-makers on the traverse, see *ibid.:* 196 – 7.
29 Our discussion of Lowe's model is relatively brief, not because of its lack of importance but rather due to the fact that most of the relevant issues have been discussed in a series of papers by Halevi, who has considered the relation between Lowe's traverse and the works of Dobb, Hicks, Kalecki and Marx. See Halevi (1983, 1984, 1992) and Halevi and Kriesler (1992). See also Harcourt (1979).
30 Kaldor, in a much neglected article (Kaldor 1938), explicitly addresses the problems caused by the specificity of factors of production.
31 See, for example, Kalecki (1971:165).
32 See Kriesler (1987:84; 1996).
33 See also Kalecki (1976:17, 20 – 2, 66).
34 See Kriesler (1994).
35 Robinson's *Collected Works* abound with such condemnation of equilibrium method. As a good example, see 'History versus equilibrium', reprinted in Robinson (1979). See also Harcourt (1981).
36 Much of this literature is discussed in Harcourt (1979).

REFERENCES

Arrow, K. and Hahn, F. (1971) *General Competitive Analysis,* Edinburgh: Oliver and Boyd.

Baldwin, R. (1988) 'Hysteresis in import prices: the beachhead effect' *American Economic Review,* 78, 773 – 85.

——(1989) 'Sunk-cost hysteresis', *NBER Working Paper* no. 2911.

Blanchard, O.J. and Summers, L.H. (1987) 'Hysteresis in unemployment', *European Economic Review,* 31, 288 – 95.

Dornbusch, R. (1976) 'Expectations and exchange rate dynamics', *Journal of Political Economy,* 84, 1161 – 74.

Dow, S. (1985) *Macroeconomic Thought,* Oxford: Blackwell.

Eatwell, J. (1982) 'Competition', in I.Bradley and M.Howard (eds) *Classical and Marxian Political Economy,* London: Macmillan, 203 – 28.

Fisher, F.M. (1983) *Disequilibrium Foundations of Equilibrium Economics,* Cambridge: Cambridge University Press.

Frank, M. and Stengos, T. (1988) 'Chaotic dynamics in economic time-series', *Journal of Economic Surveys,* 2, 103 – 33.

Gandolfo, G. (1987) 'Stability', in J.Eatwell, M.Milgate and P.Newman (eds) (1987) *The New Palgrave,* vol. 4, London: Macmillan, 460 – 4.

Goodwin, R.M. (1982) *Essays in Economic Dynamics,* London: Macmillan.

Hahn, F. (1985) *Money, Growth and Stability,* Cambridge MA: MIT Press.

Halevi, J. (1981) 'The composition of investment under conditions of non uniform changes', *Banca Nazionale del Lavoro Quarterly Review,* 213 – 32.

——(1983) 'Employment and Planning', *Social Research,* 50, 345 – 58.

——(1984) 'Lowe, Dobb and Hicks', *Eastern Economic Journal,* 10, 157 – 67.

——(1992) 'Accumulation and structural equilibrium', in D.Laibman, E.Nel and J. Halevi (eds) *Beyond the Steady State: Essays on the Revival of Growth Theory,* London: Macmillan, 264 – 88.

Halevi, J. and Kriesler, P. (1991) 'Keynes, classical economics and the surplus approach', *Review of Political Economy,* 79 – 92.

Halevi, J. and Kriesler, P. (1992) 'Introduction to the traverse', in D.Laibman, E. Nell and J.Halevi (eds) *Beyond the Steady State: Essays on the Revival of Growth Theory,* London: Macmillan, 225 – 34.

Harcourt, G.C. (1976) 'The Cambridge controversies: old ways and new horizons or dead-end?', reprinted in C.Sardoni (ed.) (1992) *On Political Economists and Modern Political Economy: Selected Essays of G.C.Harcourt,* London and New York: Routledge, 130 – 65.

——(1979) 'Non-neoclassical capital theory', reprinted in C.Sardoni (ed.) (1992) O*n Political Economists and Modern Political Economy: Selected Essays of G.C. Harcourt,* London and New York: Routledge, 166 – 79.

——(1981) 'Marshall, Sraffa and Keynes: incompatible bedfellows?', reprinted in P. Kerr (ed.) (1982) *The Social Science Imperialists: Selected Essays of G.C. Harcourt,* London: Routledge and Kegan Paul, 205 – 21.

Harris, D.J. (1988) 'On the classical theory of competition', *Cambridge Journal of Economics,* 12, 139 – 67.

Hicks, J.R. (1965) *Capital and Growth,* Oxford: Oxford University Press.

——(1973) *Capital and Time,* Oxford: Clarendon Press.

——(1985) *Methods of Dynamic Economics,* Oxford: Clarendon Press.

Kaldor, N. (1934) 'The determinateness of static equilibrium', *Review of Economic Studies,* reprinted in N.Kaldor (1960) *Essays on Value and Distribution,* London: Duckworth, 13 – 33.

——(1938) 'Stability and full employment', *Economic Journal,* reprinted in N. Kaldor (1960) *Essays on Economic Stability and Growth,* London: Duckworth, 103 – 19.

(1966) 'Causes of the slow rate of economic growth in the United Kingdom', reprinted in N.Kaldor (1978) *Further Essays on Economic Theory,* London: Duckworth, 100 – 38.

Kalecki, M. (1939) *Essays in the Theory of Economic Fluctuations,* London: Allen & Unwin.

——(1954) *Theory of Economic Dynamics,* London: Allen & Unwin.

——(1963) 'The structure of investment', reprinted in M.Kalecki (1972) *Selected Essays on the Economic Growth of the Socialist and the Mixed Economy,* Cambridge: Cambridge University Press, 102 – 11.

——(1971) *Selected Essays in the Dynamics of the Capitalist Economy,* Cambridge: Cambridge University Press.

——(1972) *Selected Essays on the Economic Growth of the Socialist and the Mixed Economy,* Cambridge: Cambridge University Press.

——(1976) *Essays on Developing Economies,* Cambridge: Cambridge University Press.

Klamer, A. (1984) *The New Classical Macroeconomics,* Kent: Wheatsheaf.

Kriesler, P. (1987) *Kalecki's Microanalysis, C*ambridge: Cambridge University Press.

——(1994) 'On Feldman's model of economic growth', in Bruce McFarlane and Peter Groenewegen (eds) *Socialist Thought in The Post Cold War Era,* Journal of Contemporary Asia Press, 27 – 38.

——(1996) 'Microfoundations: a Kaleckian perspective', in J.King (ed.) *An Alternative Macroeconomic Theory: The Kaleckian Model and Post-Keynesian Economics,* Boston: Kluwer, 55 – 72.

Lowe, A. (1975) 'Adam Smith's system of equilibrium growth', in A.S.Skinner and T.Wilson (eds) *Essays on Adam Smith,* Oxford: Clarendon Press, 415 – 25.

——(1976) *The Path of Economic Growth,* Cambridge: Cambridge University Press.

Lucas, R.E. Jr (1983) *Studies in Business-Cycle Theory,* Cambridge MA: MIT Press.

Robinson, J. (1953) 'A lecture delivered at Oxford by a Cambridge economist', reprinted in J.Robinson (1973) *Collected Economic Papers,* IV, Oxford: Blackwell, 254 – 64.

——(1953 – 4) 'The production function and the theory of capital', *Review of Economic Studies,* XXI, reprinted in J.Robinson (1975) *Collected Economic Papers,* II, 2nd edn, Oxford: Blackwell, 114 – 31.

——(1964) *The Accumulation of Capital,* 2nd edn, London: Macmillan.

——(1974) 'History versus equilibrium', reprinted in J.Robinson (1979) *Collected Economic Papers,* V, Oxford: Blackwell.

Young, A.A. (1928) 'Increasing returns and economic progress', *Economic Journal,* 38, 527 – 42.

Zeeman, E.C. (1977) *Catastrophe Theory,* Reading MA: Addison-Wesley.

23
TECHNOLOGICAL PROGRESS AND EFFECTIVE DEMAND
A Kaleckian perspective

Elizabeth Webster

INTRODUCTION

In the Kaleckian literature, there is an important emphasis on the stagnationist tendencies of modern capitalist economies. This tendency originates in the failure of effective demand to ensure adequate levels of growth in the economy. Within this context, an important potential source of demand comes from technological progress, which was, therefore, seen by many economists as providing the possibility of higher sustainable levels of consumption. This latter aspect has important implications for the sustainability of long-run growth. The question addressed in this paper is whether technological progress can generate sufficiently high levels of effective demand to overcome the stagnationist tendencies of modern economies. To answer this question we use a Kaleckian model that considers the possible effects which technological change may have on aggregate demand via lower prices, more productive resources, etc.

THE ROLE OF UNCERTAINTY

A crucial feature of Kalecki's models is his assumption that most business decisions occur in the context of behavioural uncertainty, that is non-actuarial forms of risk. The objective forces which produce investment opportunities or the possibility for selling more into the market, come and go and do not usually repeat themselves. Both Knight and Keynes argued that when this form of uncertainty exists, firms cannot base a decision on an expected value calculated from a stable probability distribution function based on a finite set of events. A probability distribution function of possible outcomes cannot be constructed because we cannot define a list ofexhaustive outcomes. Nor can we construct a set of probabilities which we hold with a lot of confidence. Where objective constraints are limited, a conjectural infinite regress may preclude a determinate solution or range of solutions (Pesaran 1987:276 – 9).

As a reasonable alternative to mathematical calculation, Keynes believed that rational firms would rely on conventions and rules of thumb. In practice this may

amount to assuming that present circumstances will continue into the future, not because they believe this is probable, but because there is no reason, on balance, why it should change in one direction and not another.[1] We base our estimates of the future states of the world, he claimed, on what we know for certain rather than speculative prophecies which we hold with little confidence or even hold as a plausible possibility, for there may be several plausible possibilities. Firms may believe that a fall in the market price for their goods will cause sales in the market to grow, but they hesitate to rush in and increase production because of the many factors which may intervene.[2] For Kalecki's model this means that firms' main systematic force in governing their decisions to produce and to invest is their most recent experience in these departments. This does not mean that irregular factors or forces stemming from events exogenous to the model will not influence these decisions. It merely argues that the simple extrapolative expectations rule is the main systematic endogenous force.

THE MODEL

In his last model, for the sake of simplicity Kalecki leaves aside lags in consumption spending, and excludes workers' saving, and changes in inventories. This leaves the basic identities (for any given short period defined as t) and short-period equilibrium condition[3]

$P \equiv C_K + S$

$W \equiv C_W$

$Y \equiv W + P \equiv C_W + C_K + S$

and short-period equilibrium condition[3]

$S = I$

so that

$$Y = C_W + C_K + I$$

where P=real aggregate profits, gross of amortisation and net of interest payments to/from the central bank; C_K=real aggregate capitalists' consumption expenditure; S=real aggregate saving, gross of amortisation (capital consumption); W=real aggregate wages; I=real aggregate investment, gross of depreciation; C_w=real aggregate workers' consumption expenditure; and Y=real output.

This gives us three sectors, the workers' consumption goods sector, the capitalists' consumption goods sector and the investment goods sector.

Capitalists' consumption is defined by a behavioural relationship

$$C_K = \lambda P + A \tag{1}$$

which in association with the identities and inserting the time subscripts gives

$$P_t = m [I_t + A(t)] \tag{2}$$

where A=the stable portion of capitalists' real consumption expenditure, which changes slowly over-time but not over the cycle; λ=capitalists' marginal propensity to consume ($0<\lambda<1$); and $m=1/(1-\lambda)<1$.

The unit period is the time taken for the entrepreneur to reconsider his/her investment decision on the basis of unexpected profits.

Because investment expenditures are the outcome of deliberate decisions and profits are not, Kalecki argued that in Equation (2), I and A are the independent variables which determine P and not vice-versa.

In any short period, we assume that production equilibrium holds, thus total output is equal to the sum of the output of the workers' consumption goods sector, the capitalists' consumption goods sector and the investment goods sector:

$$Y_t = C_{wt} + C_{Kt} + I_t$$

and therefore

$$Y_t = W_t + \lambda P_t + A(t) + f(.) \tag{3}$$

where $f(.)$ represents the equation which determines demand for investment goods.

Investment demand is considered the most volatile element in Equation (3) and Kalecki spent most of his time exploring different functional forms of investment demand. His basic method was to devise the representative firm's ex-ante investment decision equation, $f(.)$, size it up by analogy to themacroeconomic level and incorporate the macroeconomic ex-post relationship (2) to substitute out endogenous profits.

Most of his versions of $f(.)$ were designed to reflect firm's expected profitability subject to the risks associated with a high level of borrowing and debt repayments. In his 1968 paper the equation had three parts:

$eS_t=eI_t$ represented the portion of total saving in each period which formed firms' retained earnings. The higher the level of retained earnings, the greater the firms' ability to finance investment given normal prudential lending practices.

$$\left[\left(\frac{g(\Delta P_t^e, \delta)}{r} - I_t\right)\right]$$

represented the expected net present value of the proposed representative investment, where

$$\left(\frac{g(\Delta P_t^e, \delta)}{r}\right)$$

equals the present value of additional profits generated by the investment (r is the rate of discount equal to the default free rate of interest, the expected rate of

depreciation and an allowance for aversion to uncertainty), and I_t is the present value of the cost of investment.[4] P^e are the expected increase in profits which result from the proposed investment expenditure and δ is the rate of depreciation, that is, the rate at which we expect this project to siphon profits off more obsolete capital assets.

B(t) represents any additional stimulus to investment demand by those who believe that they will be the first in the field with an innovation and thus reap above expected profits.

In addition, he assumed that there was a period lag between the recognition of a need and the subsequent decision to undertake an investment. Combining these factors in an additive form gives:

$$f_t(.) = eI_{t+1} + h\left[\left(\frac{g(\Delta P^e_{t+1}, \delta P_{t+1})}{r}\right) I_{t+1}\right] + B \tag{4}$$

where h is the coefficient which translates this net present value into the investment decision.

If we let ω represent the real-wage per person employed, L (such that $\omega = W/L$) and μ represent output per worker (such that $\mu = Y/L$), then

$$Y_t = \frac{A(t) + f_t(.)}{(1-\lambda)\left(1 - \frac{\omega_t}{\mu_t}\right)}$$

Given

$$P_t \equiv Y_t - W_t$$

then from (3) the short-period equilibrium level of output can also be expressed as

$$Y_t = \frac{A(t) + f_t(.)}{(1-\lambda)\left(1 - \frac{\omega_t}{\mu_t}\right)} \tag{5}$$

From this equation we can see that a rise in the real wage rate ω will raise the output level, via the effect on consumption demand (denominator) but its effect via investment demand is unclear and depends on the functional form of *f*(.). Similarly, a rise in labour productivity will reduce output in the consumption goods sector but for the same reason as real wages, the effect via investment is also unclear.

We can substitute (4) into (5) to get

$$Y_t = \frac{A(t) + eI_{t+1} + h\left[\left(\underline{\underline{g(\Delta P^e_{t+1}, \delta P_{t+1})}}\right) - I_{t+1}\right] + B}{(1-\lambda)\left(1 - \frac{\omega_t}{\mu_t}\right)} \tag{6}$$

THE EFFECTS OF TECHNICAL CHANGE

To argue that technical change produces an endogenous rise in demand, we need to show how it leads to a rise in demand in at least one of the three sectors. In his *Studies in Economic Dynamics,* Kalecki separated the effects of technical change into three components: the price effect, the productivity of labour effect and the stimulus or innovation effect, and we will maintain this distinction. For the argument presented below, we assume that the change in technology has arrived costlessly from sources external to theeconomy, but the analysis is essentially similar for induced technical change. Furthermore, we will assume that this has resulted in a reduction in the labour costs of producing a given commodity. The case of product innovation is parallel and will not be explicitly discussed.

We assume that short-period equilibrium holds within each period, that is, firms quickly adjust their production levels to meet sales and there are no unintended accumulations of inventories. Firms seek to purchase new technologies which have just become commercially available, and will, in the process of adopting a more labour intensive technology, reduce their workforce, *ceteris paribus*. In the context of uncertainty, decision-makers tend to base their choices on hard information and facts rather than speculations based upon soft information. As such, we expect, even if product prices are reduced, that the firm will maintain their production levels at the level they otherwise would have, for the best and hardest information about a firm's short-term future sales is their current sales. Before they go to the risk of committing themselves to higher production levels, firms would expect to see some sign of production deficiency.[5] Given this, the final effect of technical change on real aggregate demand will depend on the combination of the three logically distinct effects.

The price effect

If, following the adoption of a labour saving process, firms maintain their profit margins, then prices fall (if of course prices don't fall, then there is no 'price effect' and we may proceed to the next two sections). When the technical change has occurred in the investment goods sector, the price of investment goods falls, but client firms will still choose to purchase the same real amount of investment goods which they otherwise would have, because there is no reason why their

anticipations of future sales should have changed. Firms may believe that installing new capital equipment enables them to capture market shares from their rivals because they will be able to offer cheaper goods, but we will deal with this aspect under the 'innovation effect' below, and put this issue aside here by assuming that all rival firms invest in the new cheaper capital goods in equal proportions.[6]

When technical change affects the sector producing capitalists' consumption goods, these consumers still require the same real bundle of consumer goods because real profits have not increased (both prices and thus nominal profits have fallen) and thus their real consumption demand does not rise. Similarly, a fall in both the price of wage goods and the nominal value of the wage bill will have a neutral effect on real demand for wage goods.

With real purchases unaffected, real incomes remain the same and next period's real purchases are also unchanged. Because of the different saving propensity between capitalists and workers, price reductions will only affectreal aggregate demand if the distribution of income between wages and profits is changed.

Indirectly however, Kalecki recognised that price reductions can have a marginal effect on real demands. First, a fall in nominal demands for money will, if real cash balances by banks are not reduced *pari passu,* result in a fall in the short-term and eventually long-term rates of interest (Kalecki 1943: 186). Second, Kalecki's model suggests that a fall in the price of investment goods in the case of firms, and consumption goods in the case of rentier consumers, can lead to marginally higher real retained earnings and income, *ceteris paribus,* if there are time lags between the receipt of income and expenditure. Each agent's income accrues at slightly higher price levels than those which prevail when the subsequent expenditure is made. This small real income gain should boost real expenditures. However, Kalecki did not explicitly draw the latter conclusion.[7]

The productivity of labour effect

The effect of a once-and for-all rise in the productivity of labour depends on whether it is absorbed by higher wages or higher profits per unit of output.[8] Consider first the case where nominal wages and prices do not change so the full effect is gained by profits per unit of output.

In the immediate period, firms react to the rise in productivity by reducing their labour requirements. If we assume that there is no lag between wages and workers' consumption, this fall in employment will result in a lower real wage bill and lower real spending on workers' consumption goods. To assume that firms in the workers' consumption goods sector react to a rise in labour productivity by raising or maintaining employment is to imply that they anticipate higher sales for their goods in the current period. It is unlikely that they will believe that demand for their product *will* rise in the absence of any

evidence for the proposition. Thus in terms of Equation (6), the rise in μ leads to a fall in output in the workers' consumption goods sector and in total.

Production in the investment goods sector is determined by decisions made in the previous period regarding the investment decision. While there is a higher profit share throughout the economy, the level of real output in the workers' consumption goods sector has fallen due to the lower real wage bill. The aggregate *level* of real profits remains the same and there is no additional stimulus to rentiers' consumption and investment demand. From (2), we can see that if I and A are determined prior to the start of the period, then a higher P/Y can only be accommodated by a lower Y. The final result will be lower production in the workers' consumption goods sector, the same production in the investment goods sector and capitalists' consumption goods sector, and therefore a fall in aggregate employment and real output.

We have assumed in this Kaleckian model that the production decision closely tracks sales and changes in inventories.[9] Real profits, which depend on the constant portion of capitalists' consumption and past investment decisions by capitalists, will not be affected in this current period. Assumptions regarding the lags between receipt and payment of incomes are critical to our conclusion.

It is possible nevertheless, that the fall in the costs of production alters the investment decision by making future investment projects appear more profitable. However, Kalecki argues that firms base their profit expectations largely upon current trends in profitability. The short-period effect of a rise in μ has been to produce a surplus productive capacity in firms but not a change in their aggregate level of profits. On this basis, Kalecki argues that there is no reason to expect future investment projects to appear more profitable *ceteris paribus*. That is, holding constant all the other forces which may impinge upon the desire to seek profits by investing (i.e. government policies, competition, trade-cycle events), the fall in real wages and rise in profit margins has not given firms an additional desire to invest.

If nominal wages change to match the change in labour productivity, then in the current period employment will also change proportionately, so that there will be no change in the real wage bill and, as before, no change in real profits or real output. In terms of Equation (5), both ω and μ increase in equal proportions and Y is unaffected, the only effect being a reduction in aggregate employment. Without a change in output and aggregate profits, we do not expect our investment decision equation $f(.)$ to be affected.

The innovation effect

The effect of a process innovation is less straightforward.[10] According to Kalecki (1969:442) capitalists will invest in new inventions because they expect them to be more profitable than investments which come on to the market in the current year. Although this expectation does not prove to be correct, nevertheless the

investment is still seen as successful as it enables the investor to earn profits higher than the average. In other words, had they not invested in the new invention, their profits would have been lower than they turned out to be, even though they are lower than expected.

The expected rate of decline of these profits over the life of the asset represents the economic depreciation rate. In terms of Equation (6), it is included twice as a rise in δ both increase $g(.)$ and r. The net effect from these two influences can go either way, but the faster is the rate of growth of real output (due to exogenous factors such as government expenditure and capitalists' consumption demand) the more likely it will be positive. Essentially, the effect of an exogenous rise in technical progress and thus a rise in the rate of depreciation of the value of capital assets, is to alter thetemporal distribution of profits from each vintage of capital assets. A higher rate of depreciation will cause profit flows in near-time periods to be higher, but they will be expected to decline at a faster rate over time as the asset is superseded by later vintages of greater efficiency or greater value to consumers. Consequently, the magnitude of any stimulus to the long-period equilibrium level mainly depends on the rate of discount r less depreciation (i.e. the default-free rate of interest plus a premium for aversion to uncertainty).

In terms of Kalecki's output equation (5), this innovation effect only alters the investment demand portion $f(.)$ so we are permitted to separately analyse the effect of a change of δ on $f(.)$. To do this mathematically, we must specify an equation for $g(.)$, and as an example one of Kalecki's specific equations has been chosen. Kalecki believed that the main systematic force in determining a firm's profit expectations was their recent experience with profits.[11] Modelling or even describing when and if firms believe that their past investment expectations have been realised is very difficult, in part because there are at any time several vintages of assets in place.

On this basis we may say (and we follow Kalecki here) that the firm's return to last period's investment expense rI_{t-1} includes the increment to total profits in the current period plus the extra profits the firm has retained and which otherwise would have been lost to competitors because of obsolescence or depreciation of its existing capital assets. Thus the extra profits are an implied return. If rI_{t-1} is the level of profits the firms require in period t in order to meet their minimum ex-ante requirements, then we can say that investment expectations are fully met in period t when

$$\Delta P_t + \delta P_t = rI_{t-1}[12]$$

Kalecki puts this relationship in present-value terms and assumes that current realisation of past expectations form the basis for current expectations. The present value to a stream of future profits which decline at a geometric rate of δ should be discounted at the rate $(i+x)$, where i is the default-free rate of interest

(the central bank cash rate) and x is the premium to compensate firms for their aversion to uncertainty. The present value is therefore

$$\frac{(\Delta P^{e}_{t-\tau},+\delta P^{e}_{t-\tau})}{r}$$

If we assume that $P^{e}=P$, then the present value of future expected profits,

$$\frac{(\Delta P^{e}_{t-\tau},+\delta P^{e}_{t-\tau})}{r}$$

A long-period equilibrium equation can then be derived as a particular integral of the short-period macrodynamic equation. In his 1968 paper, he defines his short-period investment demand equation as:

$$I_{t+\tau}=\left(e-h+\frac{hm}{r}\delta\right)I_t+\frac{hm}{r}n\Delta I_t+\frac{hm}{r}\delta A(t)+\frac{hm}{r}n\Delta A(t)+B \qquad (7)$$

Where r=the gross required rate of profit, $r>0$ (it is the rate of gross profit firms require before they are prepared to invest); h=measure of entrepreneurs confidence ($h>0$); e=portion of profits retained by firms ($0<e<1$); δ=rate of obsolescence due to technical progress ($\delta>0$); n=the pro rata share of additional profits captured by the new investment in the absence of technological progress ($0<n<1$); and B=the effect innovation has on entrepreneurs who expect to capture rents from being the first in the field to innovate.

Kalecki simplifies (7) to

$$I^{lp}_{t}=\frac{ce^{\beta t}}{e^{\beta t}-a-b\beta}$$

Where τ=the lag between the investment decision and actual expenditure; $a=e-h+(hm\delta)/r<1$, to assure oscillations rather than monotonic growth; $b=(hmn)/r$. If $b<1$, oscillations dampen, $b>1$ oscillations explode and $b=1$ oscillations are uniform; and $F(t)=(hm\delta/r)A(t)+(hmn/r)\Delta A(t)+B(t)$.

And subsequently his particular integral is

$$I^{lp}_{t}=\frac{ce^{\beta t}}{e^{\beta t}-a-b\beta} \qquad (8)[13]$$

$$\text{s. t. } e^{\beta\tau}-a-b\beta>0$$

where I^{lp}=trend level or long-period equilibrium level of gross investment expenditure.

We use the term long-period equilibrium in the sense of a frictionless pendulum, which always passes through the same point on its swing.[14] It is neither a position the economy will tend towards if it is left to settle without exogenous interference, nor an average position from which real positions deviate randomly. Our long-period equilibrium indicates the average level of economic activity over the whole trade cycle, given the level of autonomous expenditure and for given values of our behavioural coefficients.

The effect of technical change is measured by the effect on the level or rate of growth of the long-period equation (8). Consider a rise in the rate of economic depreciation due to say, an exogenous change in science and technology. Two opposing influences will occur: a higher rate of product improvement, process efficiency, etc. will, by raising δ, siphon profits off old capital assets to new investments, faster than otherwise. This will stimulate investment demand; a higher expected δ will raise the required rate of profit, $r=i+x+\delta$. Thus entrepreneurs will demand a higher expected gross rate of return before they are prepared to invest. The higher is r, the fewer will be the value of investment opportunities which will be assessed as viable *ceteris paribus*.

Mathematically the net effect depends on the influences on a and $F(t)$. Specifically, given

$$a = e - h + \frac{hm\delta}{(i + x + \delta)}$$

$$i + x + \delta > 0.^{15}$$

Then

$$\frac{da}{d\delta} = \frac{hm(i + x)}{(i + x + \delta)^2} > 0$$

Similarly, given

$$F(t) = \frac{hm\delta}{(i + x + \delta)} A(t) + \frac{hmn}{(i + x + \delta)} \Delta A(t) + B$$

$$i + x + \delta > 0.^{16}$$

then

$$\frac{dF(t)}{d\delta} = \frac{hm\{A(t)(i + x) - n\Delta A(t)\}}{(i + x + \delta)^2} > 0$$

as $n\Delta A(t)$ is very small and can be ignored.

We may then conclude that according to Equation (7) a once-and-for-all rise in δ will cause both a once-and-for-all-rise in a and a once-and-for-all rise in $F(t)$ and this will lead to a once-and-for-all rise in I^{lp}. Essentially the greater success of a more rapid rate of new ideas and innovations in stealingsales from past vintages of investment more than offsets the faster decline of these profits over time.

In his 1943 *Studies in Economic Dynamics,* Kalecki discusses the two opposing effects, given the above, on the incentive to invest of a change in the depreciation rate. But he leaves open the possibility of either forces dominating (Kalecki 1943:187 – 90). However, our reinterpretation of his 1968 model, especially the inclusion of δ in r and the assumed sign of the coefficients, allows us to conclude that the existence of technical progress by raising δ will have an overall positive effect on the incentive to invest. Nonetheless, a constant but positive rate of technical progress and innovation does not induce a positive *trend* rate of growth but does raise the *level* of the long-period equilibrium around which real investment and income fluctuate.

An increase in δ will only affect the trend rate of growth if b is affected. However, the effect of a change in δ on b is less clear-cut, b represents the weighted rates of growth between $A(t)$ and B (the effect of $\Delta A(t)$ being insignificant). If the rate of growth of $A(t)$ exceeds that of B which is probable on economic grounds, then a rise in δ which will increase the relative weighting towards $A(t)$, will raise the estimated b. However, if the rate of growth of both $A(t)$ and B are zero, then a change in δ will not affect the rate of growth of I^{lp}, which will remain zero.

CONCLUSION

Kalecki's models incorporate an expenditure-income flow view of the economy. Incomes are generated by expenditure, and the main systematic force determining expenditures are incomes or changes to incomes. Irregular forces such as those discussed above under the innovation effect, do exist, but there is no necessary reason for them to grow over time. Firms will not, on average, increase production until they have solid evidence of an increase in demand, but without this increase in production and incomes there is little reason to expect expenditure to increase. There may be an increase in the level of buoyant and confident entrepreneurs who, on some basis, be it the ownership of a technical innovation or a good idea, will increase their production, but unless the percentage of enthusiasts increases over time, this will merely create a higher level of production and not a continual growth in production.

The effect of many changes to the macro economy depends importantly on the temporal ordering of events and outcomes. In our case, the fact that the increase in labour productivity leads in the first instance to a fall in employment is critical because consumption demand changes without a lag while investment is delayed

in time. This assumption may not be always valid, and we cannot therefore claim that Kalecki's reasoning is entirelygeneral and applies in all circumstances. However it seems reasonable to assume that it applies if we also accept the Shackle-Keynes simplified decision-making rules under economic uncertainty. While not all economists subscribe to a Kaleckian view of macroeconomics and may find the analysis presented above alien, it nevertheless remains the responsibility of those who purport that technical change does go hand-in-hand with an expansion of aggregate demand, to provide a convincing process by which this is achieved.

ACKNOWLEDGEMENTS

I would like to thank Geoff Harcourt, Robert Dixon, Malcolm Sawyer and Geoff Meeks for comments on a more primitive version of these arguments. Usual exonerations to hapless readers apply.

NOTES

1 Keynes (1937). Meeks (1991) argues that this form of decision-making is quite rational, in the sense of reasonable, in these circumstances.
2 For example, consumers may respond slower than expected to price changes, rival product prices may fall also, or there may be unanticipated changes in the macro economy.
3 Short-period equilibrium occurs when production expectations are realised so there is not unintended accumulation of stocks.
4 A simple extrapolate expectations model was assumed for profit expectations.
5 This may mean that firms have innovated in the hope that demand will rise due to the fall in costs of production, but that they wait to see the effects on demand before they hire more labour to produce more products.
6 Assuming that the investment goods price fall is passed on to consumers, there may be a case for expecting that the size of the market will grow. If we believe that firms hold this expectation with a solid conviction, then, they will demand more of the cheaper investment goods. However, this view is unlikely if we accept Keynes' notion that under uncertainty, firms base their business decisions more on hard information than speculations about what could be. That is, they wait and see if demand picks up before they commit themselves to higher production levels.
7 Nevertheless, both these effects are excluded if, as it is increasingly alleged, central bank behaviour results in an endogenous money supply (see Lavoie 1992).
8 We hold prices constant in this section, having dealt with the price effect above.
9 This follows from the assumption that short-period equilibrium is maintained within the period.
10 Bear in mind that we are speaking of a process innovation only to reduce confusion to the reader. A parallel case can be made for a product innovation, but not here.
11 As a modeller we have to ask how and to what extent firms can and do decide that an increase in their profits was due to one asset over another, or due to their own

investments rather than general economy-wide circumstances. How do they cope with synergy effects? Do they discount profits which are due to general trade-cycle effects rather than what they feel is the 'intrinsic' merit of their own investments? One suspects that there is a degree of uncertainty and ambiguity in business over how to apportion profits. Furthermore, there are good reasons for arguing that even if your profitability came from trade-cycle effects, to capture (avoid) further trade-cycle effects, one should continue to (desist from) investing.

12 This automatically assumes that investment expectations previous to I_{t-1} have been fulfilled, and thus any discrepancies in expectations in period t are attributed to the investment I_{t-1}.

13 There is a typographical error in Kalecki's *Collected Works* (1968:444), and readers should refer to the original *Economic Journal* article or 1971 book.

14 Harcourt (1981:251).

15 a approaches $e-r+rm>0$ asymptotically, as $\delta\rightarrow\infty$.

16 $F(t)$ approaches $rmA(t)+B(t)>0$ asymptotically, as $\delta\rightarrow\infty$.

REFERENCES

Harcourt, G.C. (1981) 'Marshall, Sraffa and Keynes: incompatible bedfellows?', in C.Sardoni (ed.) (1992) *On Political Economists and Modern Political Economy: Selected Essays of G.C.Harcourt,* London and New York: Routledge, 250 – 64.

Kalecki, M. (1934) 'Three systems', in J.Osiatynski (ed.) (1990) *Collected Works of Michal Kalecki: Volume I, Business Cycles and Full Employment,* Oxford: Clarendon Press, 201 – 19.

——(1943) 'Studies in economic dynamics', in J.Osiatynski (ed.) (1991) *Collected Works of Michal Kalecki: Volume II, Capitalism Economic Dynamics,* Oxford: Clarendon Press, 117 – 90.

——(1968) 'Trend and the business cycle', in J.Osiatynski (ed.) (1991) *Collected Works of Michal Kalecki: Volume II, Capitalism Economic Dynamics,* Oxford: Clarendon Press, 435 – 50.

Keynes, J.M. (1937) 'The general theory of employment', *Quarterly Journal of Economics,* 51 (February): 209 – 23.

Lavoie, M. (1992) *Foundations of Post-Keynesian Economic Analysis,* Aldershot: Edward Edgar.

Meeks, J.G. T. (1991) 'Keynes on the rationality of decision procedures under uncertainty: the investment decision', in J.G.T.Meeks (ed.) *Thoughtful Economic Man,* Cambridge: Cambridge University Press, 126 – 60.

Pesaran, M.H. (1987) *The Limits to Rational Expectations,* Oxford: Blackwell.

Shackle, G.L.S. (1959) 'Time and thought', in J.L.Ford (ed.) (1990) *Time, Expectations and Uncertainty in Economics: Selected Essays of G.L.S.Shackle,* Aldershot: Edward Edgar, 14 – 27.

Trevithick, J.A. (1992) *Involuntary Unemployment,* New York: Harvester/Wheatsheaf.

Worswick, G.D.N. (1977) 'Kalecki at Oxford', *Oxford Bulletin of Economics and Statistics,* special issue 39 (1):19 – 29. Oxford: Blackwell.

24
LIGHTHOUSES, TELEVISION AND THE THEORY OF PUBLIC GOODS

Diana Barrowclough

> *Television?* The word is half Latin and half Greek. No good can come of it.[1]

INTRODUCTION

Recent reforms in broadcasting policy are challenging long-held views about the characteristics of 'public goods' and the appropriate role that these imply for the state. Broadcasting's traditional characteristics of non-excludability, non-rivalry, and positive and negative externalities have long been cited to justify an important role for the state. State intervention was needed to help achieve society's mutually desired goals, as the problems of ill-defined property rights, opportunism and free-riding would undermine our ability to coordinate ourselves voluntarily.

This is an old story in economics, as the sharing of collective goods is an intrinsic feature of social life which has long exercised economists from Hume, Smith and Mill onwards. How does society organise itself to produce those goods and services which are desired by all, but which will be under-provided, if provided at all, when left to our own opportunistic and short-sighted devices? The earliest economists used the analogy of the lighthouse to explore these issues, and reached the conclusion above: that the rational choice for an individual facing conditions of non-excludability and non-rivalry leads to collective inaction, thus justifying state intervention to ensure the socially optimal outcome.

When broadcasting was invented at the turn of the century it was swiftly subsumed into the same metaphor, and the same conclusion reached. Broadcasting also was non-rival and non-excludable; moreover it was strategically, socially and politically too important to be left to the impersonal forces of the market. This view was clearly manifested in policy, to the extent that a strong state role emerged in broadcasting throughout theworld. The precise nature of this role varied from country to country—running the gamut of possibilities from minimal 'night-watchman' state regulation of private

enterprise, to a fully integrated state funded, owned and operated broadcasting service. Throughout however, a strong state role was the norm and not the exception.

Recently, however, an extraordinary wave of reform has swept through public broadcasting throughout the world, assailing the traditional state role from many quarters. State ownership has been questioned, as has the state's role in collecting what is in many countries the last targeted, virtually universally levied broadcasting tax. Quasi-markets for broadcasting have been introduced in those countries which want both reform and continued loyalty to traditional public-service broadcasting ideals; while other countries have doubted whether broadcasting contains any 'public good' aspect at all.

This is by no means unique to broadcasting—rather it echoes familiar themes heard throughout all aspects of the public economy. As the public sector accounts for between 40 – 50 per cent of GDP in many western countries, virtually no service remains undisturbed by the quest for increased efficiency, accountability, and perceived value for money, quite apart from the effect of wider 'neo-Liberal' calls for a more limited state. These challenges fall into four broad themes:

1 Public provision cannot be efficient, because of the 'preference revelation problem'.
2 'Government failure' is pervasive and worse than the 'market failure' it aimed to remedy.
3 Public provision is not the only alternative—markets or voluntary associations can, and in the past did, provide goods with public characteristics.
4 Public goods do not really exist anyway, at least in the 'pure' case envisaged.

Reflecting all these challenges, the lighthouse metaphor itself has been stripped bare—as Coase (1988) and Peacocks' revised histories reinforce wider claims that markets can, and did, successfully provide 'public goods.' This reflects the general challenges to the state's role in all public services, but is especially interesting for broadcasting in particular, given the importance of Peacock's 'watershed' 1986 report on broadcasting and Coase's early forays into broadcasting economics.

The next section therefore very briefly distils the 'conventional wisdom' on public goods theory over the last two centuries, showing its traditional eighteenth- and nineteenth-century affiliation to lighthouses, and its twentieth-century extension to broadcasting. The third section introduces recent challenges to this view, while the final section concludes by highlightingsome of the main questions which these challenges pose for public (and broadcasting) economists as we enter the next century.

THE CONVENTIONAL WISDOM

Market failure and the problem of exclusion

For most of the decades since broadcasting's invention, the following core 'stylised facts' about broadcasting have been accepted (by 'consumers' we mean viewer/listener audiences; by 'providers' we mean programme producers, both private and public sector, and/or broadcasting operators such as the BBC):

- Broadcasting was traditionally non-excludable, in that historically broadcasters lacked the technical means to exclude any listeners/viewers from the consumption of broadcast programmes (given the initial purchase of a receiving set).
- Broadcasting is non-rival, as consumption by one listener/viewer does not reduce the consumption benefits available to others.
- Unlike other industries there is no direct link between final consumer and provider. The price mechanism fails, so consumers (audiences) cannot signal the intensity of their desires through revealed willingness to pay for different services. The only signal consumers can send is that of audience size—the number of people switched 'on' rather than 'off'.

These features alone would have been sufficient to argue that, left to its own devices, society would be unable to coordinate itself voluntarily to produce the quantity and quality of broadcasting desired. The technological features of non-excludability and non-rivalry, stemming from ill-defined and unenforceable property rights, prompts us to assume that the market would fail. Opportunism, coupled with the potential for free-riding, would undermine society's ability to reach the universally desired cooperative solution.

The implications these features present for public policy have been identified from the early 1700s. The 'Prisoner's Dilemma' treatment shown below is essentially just a modern presentation of Hume's 1739 observation that two neighbours could agree to drain a shared meadow, but a larger communal ownership would fail to coordinate itself. Even though the drained meadow was desired by all, collective inaction was the inevitable outcome, given the features of non-excludability and non-rivalry, and the behavioural motivations of opportunism and free riding (see Table 24.1).

Similarly, Smith's much quoted 1776 reiteration that ill-defined and non-enforceable property rights require the intervention of some higher authority to ensure the provision of 'those institutions and public works which, though they may be highly advantageous to a great society, are, however, of such a nature that the profit could never repay the expense to any individual'.

Table 24.1 Modern depiction of Hume 1739

Consumer A/Consumer	*does not free-ride*	*free-rides*
does not free-ride	1 [10,10]	4 [6,12]
free-rides	2 [12,6]	3 [0,0]

Thus intervention by a higher authority was required for society to achieve the advantageous public works embodied in cell 1 of Table 24.1.

This issue is now so familiar to modern economists that first-year undergraduates are expected to know at a glance what the game above represents, even if they are shown only the respective pay-offs and not the description of the various 'free-ride' or 'not free-ride' strategies. The important point to note is that this Prisoner's Dilemma treatment reflects a lineage virtually unbroken over 300 years. It is only in the last decade or so that the more radical challenges posed have questioned the game's assumptions or its payoffs.

In the 300 years between Hume and public economists of today, many writers echoed these thoughts, and introduced the lighthouse example which was to appear again and again in the centuries that followed until broadcasting usurped its place. J.S.Mill's *Principles of Political Economy* invokes the lighthouse, and prefigures the Public Broadcasting Fee tax:

> it is a proper office of government to build and maintain light-houses…for since it is impossible that the ships at sea which are benefited by a lighthouse, should be made to pay a toll on the occasion of its use, no one would build lighthouses from motives of personal interest, unless indemnified and rewarded from a compulsory levy made by the state.
>
> (Mill 1965:968)

Certainly the experience of broadcasting has been that some form of third-party payment had to be organised, given the difficulties of extracting payment from consumers at source. In the public sector, Mill's 'compulsory levy' was charged to all owners of a television set, and then used to fund programme production and broadcasting; so programming could then be provided 'free' at source to (non-excludable) consumer audiences. Thus compulsory taxation takes society to cell 1 in the prisoners' dilemma above.

Of course commercial broadcasting has also emerged, so we must ask how it came about given the non-excludability and non-rivalry features. There was historically still no direct link between consumer (viewer) andprovider, as one would expect given non-excludability, however innovative ways of extracting payment from excludable, intermediate-consumers developed (often before the government-sponsored system). Broadcast operators can extract payment from intermediate consumers—advertisers—who could otherwise be excluded from

access to the audiences which comprise their target markets. Advertisers pay for access to the airwaves, and programmes are offered 'free' to consumers as a gift, to encourage them to form the audience which would then consume the advertisements.

Thus we can see that left to its own devices, society would find some way of enforcing some property rights and extracting payment for services, even if final consumers are non-excludable. However, what would be produced under these circumstances? Is it the same 'public good' that was envisaged in the Hume and Prisoner's Dilemma examples above? In modern language, are the pay-off matrixes the same as in the cells above? As drawn, the matrix could easily represent public broadcasting, where all consumers have the incentive to free-ride, each individual is best off if they free-ride and everyone else pays, and society as a whole is best off if everyone pays. However in a commercial situation the pay-offs are different. Advertisers receive utility from paying for broadcasting, as their payment is an investment in marketing, so the free-riding option for them would be a commercial mistake as it precludes access to their market. We must also ask if the product supplied to 'free-riding' consumers is as highly valued by them, given that it now includes advertisements, and possibly different sorts of programmes; i.e. does cell 2 still have the pay-off [12, 6] or should it be something more like [10, 12] reflecting the payer's positive utility from paying, and the different product provided?

Other sources of market failure

Other market failures include externalities, missing markets, imperfect information and myopic consumption decisions, and imperfect competitive structure. Again, traditional public goods theory, developed initially through the metaphor of the lighthouse, throws further light on these questions. Theories of externalities, missing markets and imperfect competition show that unregulated commercial providers of such goods would under-provide, or provide something less valued, thereby denying consumers potential surplus. In the case of broadcasting, a second tier of widely accepted 'stylised facts' lends support to this:

> Unregulated profit-maximising output for broadcasters would therefore be low-cost programmes which can capture mass audiences (who may value programming at very low rate), rather than small audiences (who may value programming very highly). This occurs because:

- the broadcast industry exhibits increasing returns to scale, in that an increase in the numbers of consumers incurs no (or very little) marginal costs to producers

- as consumers can not signal intensity of preferences, and as broadcasters have little incentive to calculate them anyway, commercial broadcasters aim instead for audience maximisation.

This has lead to criticisms of 'lowest common denomination' programming, 'ghettoisation' (putting smaller-audience programmes on in graveyard hours), and lack of differentiation in programmes.

Hotelling's model of the 'ice-cream sellers on the beach', for example, adds much insight into the common complaint of lack of diversity in programming. Dixit and Stiglitz's investigation of the optimal number of varieties in the market also reflects these issues. The point is, briefly, that an unregulated, commercial broadcasting system would lead to a significant loss of potential consumer surplus. In large part this stems from the failure of the price system to adequately reflect intensities of preferences.

Moreover, the failure of the price system is further exacerbated by the presence of consumption and production externalities, both negative and positive, which further distort recognition of true costs of resources, and consumers' valuations.

Pigou had pointed this out in the 1930s when he used the lighthouse example to discuss what he called 'uncompensated services' where 'marginal net product falls short of marginal social net product, because incidental services are performed to third parties from whom it is technically difficult to extract payment' (Pigou 1932:183 – 4).

Samuelson reinforced this argument, and introduced merit goods as well, when he argued that lighthouses exhibited 'external-economy divergence between private and social advantage'. This, he argued, meant it was necessary that 'government provides certain indispensable public services without which community life would be unthinkable and which by their nature cannot appropriately be left to private enterprise' (Samuelson 1964:45).

The earliest arguments about externalities in broadcasting included strategic and defence effects, which necessitated the state's firm control over the infant technology.[2] From the 1920s onwards interest in externalities was more related to broadcasting's ability to educate and extend (as well as entertain). Many such arguments about broadcasting's externalities remain strong today, although the days of state monopoly seem very distant. Much attention in the 1990s has focused upon consumption externalities, although production externalities (such as training effects) have also been identified. It is often suggested that given television's influence upon viewers' social habits, beliefs and interests, broadcasters should consider the preferences of those who are affected indirectly, as well as the preferences of those who choose to consume broadcasting directly. Programmes condoning racism orviolence, for example, are seen to produce negative social externalities; while educational or edifying programmes produce positive ones. (Here one can see that broadcasting is

different from the lighthouse example, where the issue is largely one of whether the light is 'on' or 'off'.)

The argument is that although it is extremely difficult to take all these indirect effects into account, a public service broadcaster is more likely to be responsive to these social effects than would be a commercial one. The BBC's first Director General, Lord Reith, embodied this approach, as his philosophy was that people be introduced to new forms of entertainment and education that they would not have consciously chosen themselves, in the hope this would help extend both their human capital and their enjoyment of life. Reith famously claimed that 'few [listeners] know what they want, and very few want what they need' (in Briggs 1979:238).

Reith's philosophy was lauded for many years, and it is only relatively recently that less emotive and paternalistic public goods theories are preferred. Arrow-inspired arguments about imperfect information and experience goods, or that consumers may be myopic about their own long-term interests, are more favourably received. For example:

- Broadcasting is an 'experience good' so that viewers or listeners cannot know how much they would value consumption of a programme until after they have consumed it.
- Moreover consumers cannot learn, as past consumption experiences are not an adequate indication to potential future valuations, because novelty is an essential feature of the desire to consume further broadcasting.
- It is not clear how consumer taste and preferences are developed in the first place, given that consumers can only show (limited) response within the choices that they are offered.

The position in the 1990s is well summarised by Graham:

> The point being made here is not that television may have great power for good or evil over society as a whole, but that television has the capacity either to cramp or enrich the knowledge, experience and imagination of individuals…if all television is elicited by the market, there is a very real danger that consumers will under-invest in the development of their own tastes, their own experiences and their own capacity to comprehend. This is not because consumers are stupid, but because it is only in retrospect that the benefits of such investment become apparent.
>
> (Graham 1992:174)

Finally, industry structure arguments are commonly used to show theneed for strict government regulation to ensure the public goods aspects of broadcasting are maintained: the industry structure internationally is typically highly horizontally and vertically integrated; and dominated by a small number of

'core' companies. This tends to reinforce, or is reinforced by, the industry's inherent incentive to maximise market share.

Thus it is commonly argued that the factors of non-rivalry, non-excludability, externalities and imperfect information are exacerbated by a market structure which can be characterised by horizontal and vertical integration, information asymmetries, and high transactions costs. Those prices which do emerge appear unlikely to reflect society's 'true' shadow prices. Harcourt (1992) has colourfully described the conditions under which prices can properly reflect the total opportunity cost of resources, being influenced only by relative scarcity or abundance and consumer preferences, and undistorted by any other factors. Power must be evenly diffused and all market participants price-takers; production and consumption processes must be quickly equilibrating, and without cumulative stock-effects. Using the analogy of the wolf pack, the market must be such that 'lone wolves' who run ahead or behind are forced to return to the pack. In broadcasting, it is usually argued that these forces are weak, and imperfect.

Many of these arguments can be subsumed into the question of whether the consumer is 'sovereign' or not. This takes an especially emotive appeal given that we are asking whether consumers are sovereign within their own sitting rooms, let alone within their broader societies. As Adam Smith has argued, 'Consumption is the sole end purpose of all production and the interest of the producer ought to be attended to only so far as it may be necessary for promoting that of the consumer' (in Heilbroner 1986:284).

All the aspects of these arguments above can be subsumed within Samuelson's well known treatment of public goods (Samuelson 1954). Public goods are different from private goods because of their features of non-excludability; i.e. consumption of private goods is $C^i_y=C^a_y-C^b_y$, while consumption of non-divisible public goods is $C^i_g=C^a_g+C^b_g$ (where superscripts a and b stand for individuals, i for the aggregated society, and subscripts for the goods). The optimal provision of private goods therefore would be where:

$$MRS^a_{xy} = MRS^b_{xy} = MRT_{xy}$$

whereas by comparison the optimal public goods level is where:

$$MRS^a_{gx} = MRT_{gx} - MRS^b_{gx}$$

For private goods, price signals are able to operate effectively to ensure equality between MRS and MRT, but in the public sector intervention by 'higher authority' is needed.

The conventional conclusion

Thus for all of these reasons described above, the broadcasting sectors in most nations have been marked by a strong presence of the state. Rationales for the

state's role have changed over time—from strategic defence of the realm (early 1900s) to defence of powerless and ill-informed consumers (late 1900s)—and an extraordinary variety of organisational and institutional arrangements have emerged, all with the state in some central role. Throughout the world almost every form of organisational structure has been tried at one stage or another, from

1 the state as minimal 'night-watchman' regulator, simply allocating spectrum frequencies; to
2 more interventionist state regulation posing quality and content standards for private enterprise; and
3 state ownership, funding and operation of broadcasting services.

CHALLENGES TO THE CONVENTIONAL WISDOM

We can use the Samuelson synthesis above to analyse this decade's fundamental challenges to the conventional wisdom. In broadcasting, as in every public sector generally, each component of the equation above has been doubted: there is no confidence that MRS^{i}_{gx} can be identified; there is suspicion that government intervention causes an inefficient MRT_{gx}; and moreover, there is doubt whether public goods, as defined by $C_g=C^{a}_{g}+ C^{b}_{g}$, even exist in the first place.

The problem of preference revelation is crucial. If we cannot even know MRS^{i}_{gx} how can we hope that central coordination can achieve it? Later public economists added another dimension to this problem—by claiming that even if we did know *MRS,* government action could not actually produce an outcome which met it. Public choice economists showed that interest groups, bureaucrats, and other 'rent seekers' may cause government policy to diverge from that preferred by the representative *MRS* embodied in the median voter (Buchanan, Niskanen, Downs, etc.). And others claimed that even if the 'right' policy is chosen, pervasive 'government failure' means that it will not be efficiently enacted. Principal/agent theory has been harnessed to public economics, to design organisational and incentive structures that aim to minimise divergence of interest and make public institutions more responsive to the wishes of consumer/voters.

Encompassing these threats to the conventional wisdom is a growing doubt that public goods really exist at all in the way described by the lighthouse metaphor. Are there really any goods which are truly non-excludable and non-rival? Increasingly, the definition of 'publicness' is being dissectedand broken down to its smallest constituent part, in both economic theory and its policy application. Just as the theory of the firm has refined its analysis to Coase's single transaction, so too does public economics search for the single, minimal island of 'publicness' embedded within the wider sea of private attributes.

Hand-in-hand (it is not clear which is leading which) with this reappraisal of the state's role has been an important reappraisal of the metaphor of the

lighthouse. Coase and Peacock have shown that from the sixteenth to the nineteenth centuries private sector entrepreneurs successfully provided lighthouse services. The Crown did have a crucial role but this was confined to being a regulator: the Crown allocated monopoly property rights, to ensure no confusing duplication of signals in one area; it enforced externalities, such as important navigational landmarks; and it regulated prices, to prevent exploitation of a captive market. Lighthouse dues were collected when ships docked in harbour, and this was enforced as customs clearance was denied until the dues were paid. A sophisticated payment system operated, with dues varying according to ship size and port of origin, and certain kinds of ships (usually small, and local) were exempt. Thus they argued, economists' use of the metaphor over the centuries has been misguided.

Conventional thinking about public goods and lighthouses has therefore been questioned at all levels: in theory, in policy and in the commonly used lighthouse metaphor. Keynes' famous observation of the power of ideas seems particularly apt, as the 1990s becomes a decade increasingly marked by the impetus to reform as many former state functions as possible.

The effect on public broadcasting

These influences have been manifested in reforms which have redefined property rights and introduced degrees of exclusion, in an attempt to capture the perceived benefits of privateness, markets and competition.[3] In terms of the Samuelson equation earlier, the reforms have aimed to help reveal consumers' true MRS^{i}_{gx} and to improve the efficient use of resources as embodied in MRT_{gx}. Two categories of reform have emerged:

1 Direct exclusion of final consumers, by introducing a direct market. Technological advances now mean that the former system of universal access and universal payment by compulsory taxation can be replaced with selective access and selective payment by subscription.
2 Indirect exclusion of broadcasting providers from the public purse, through the establishment of a quasi-market between finance and provision. Funding is still through public revenue, collected by a virtually universal tax, and broadcasting services are still offered 'free' to consumers, but a competitive market is introduced between providers of broadcaster services.

Both reforms reflect changes to the definition of broadcasting's 'publicness', and its sources.

Exclusion—solving the problem of 'revealed preferences'?

The problem of inducing society to reveal preferences has led economists to expend a great deal of effort to estimate shadow prices for broadcasting, including contingency and hedonic pricing, which attempted to isolate the value of otherwise 'intangible' effects; Clarke-Groves taxes which attempted to minimise survey respondents' incentives to misrepresentation; Vickrey price auctions; and Willingness To Pay surveys. Typically results were very sensitive to model specification, and it was difficult to draw any clear conclusions: for example, one UK study estimating WTP found that 10 per cent of the survey population would pay £120 per annum to consume BBC1 (NOP 1986); another study found 15 per cent would pay that amount (Booz *et al* 1987); while another argued that 75 per cent would pay it (Ehrenberg and Mills 1990). As the Peacock Report (1986) diplomatically put it, the results were 'awkwardly inconsistent, and we cannot place much reliance upon them'. Other attempts to calculate MRS through revealed choices were similarly limited: insights from Tiebout spatial choice models, for example, are limited in that consumers can only choose from what they are offered.

Such widely diverging responses makes it difficult to argue that government can estimate MRS, so it was not surprising when the Peacock Report (1986) strongly recommended immediately replacing the compulsory broadcasting tax with a voluntary subscription charge (reform 1 above). There would no longer be the 'universal access' which had long justified the universal tax, as access to broadcasting could now be 'selective.' Audiences which highly valued specialist programming could signal this, in their willingness to pay higher amounts for particular programme types. Pay-television had already existed in the commercial sector, but this was the first time it was recommended for the public sector too, and the importance of the Peacock Report cannot be overestimated. For many countries the UK used to represent the 'public service ideal' towards which most nations' broadcasters aspired, and the Peacock Report was digested by interested broadcasting analysts throughout the world.

New Zealand for example, was one of the first countries to adopt Peacock recommendations. New Zealand has yet to exclude consumers by use of subscription funding, as recommended by Peacock, but it has followed the second reform structure described above, which excludes broadcasting providers from automatic access to the public purse.

New Zealand's 1990 Broadcasting Act reflected wider moves towards internal markets throughout the world's public sectors. The Act separates broadcasting purchasers from providers, and requires that providers should compete, using prices and contracts. As above, there already was this splitbetween financiers and providers in the commercial sector, where advertisers were the financiers, and providers comprised both privately owned and state-owned broadcasters. Consumers remain the viewing audience, who are still offered programming 'free' in return for creating the audience required by advertisers. What the Act

does now is to create a second competitive market, this time between public finance (the broadcasting fee revenue) and providers (again, both privately owned and state-owned broadcasters). The financier in this case is the quango, which collects and disseminates the public broadcasting fee. So the New Zealand reforms are 'quasi-market' in that

1 a new market has been created, as a number of diverse providers compete for access to the public purse, and
2 it is a quasi one, in that there is no direct access between.taxpayers themselves and providers, but rather third-party access via the quango.

Exclusion is still therefore indirect, via providers rather than via consumers. Direct exclusion of final consumers is an option which New Zealand has not yet officially explored, but it may well do so, especially given the recent development in the UK, where a new Royal Charter (Department of National Heritage 1996) has opened the door for the BBC to raise funds by subscription, 'pay per view', and advertising, or 'such other form of funding or payment…as may from time to time be approved by our Secretary of State.'

Given that exclusion is now possible, does this change conventional wisdom about public broadcasting? Public goods theory has typically started by considering the nature of the underlying good and service, to determine whether or not it was technically excludable. Now we must go the other way, considering how the change to excludability might effect the nature of the good. And the debate becomes even subtler when we ask the normative question whether it should be excludable.

The search for the source of 'publicness'—do 'public goods' really exist?

Buchanan has argued that separating out the essential characteristics of rivalry and excludability reveals that very few services actually fit into the 'pure public goods' case of cell D in Table 24.2 below. Coase and Peacock's reappraisal of the lighthouse raised doubts about whether they ever deserved to be included in cell D, and today with new satellite and electronics technology they can fit into the excludable but non-rival cell C or may indeed be so technologically altered that they fall into the private goods case of cell A (for example, private satellite navigation systems rather than the public service of the beam of light).

Buchanan argues that cells C (club goods) and B (the commons) are the only interesting cases these days, unless one wants to argue that some good or service be considered non-excludable for legal or ethical, not technological, reasons. His theory of clubs shows that goods which are non-rival but excludable can be efficiently provided through markets, through consumption-sharing arrangements. Optimal club size for example, maximises the difference between

Table 24.2 Rivalry and excludability

Characteristics	*Excludable*	*Non-excludable*
Rival	pure private good (a)	mixed good (b)
Non-Rival	mixed good (c)	pure public good (d)

benefits (which first rises and then falls as membership size increases to the point of congestion) and average costs (which fall as membership increases).

Peacock has also argued that non-excludability may be a transitory feature, and that technological advances will enable exclusion if there is no reason to expect state intervention. Entrepreneurs will 'seek out ways and means…of rough and ready fashion' to appropriate economic rent. These ways and means might include process innovations to make the consumer excludable; the invention of entirely new products; or an altering of consumers' tastes and preferences away from the non-excludable and towards the excludable. That is, left to their own devices the market will produce 'public' goods, but of type C or A.

In the late 1990s broadcasting is certainly potentially in Buchanan's mixed good category C, and not the pure D category of the traditional lighthouse metaphor. First, as discussed earlier, exclusion of intermediate parties is possible —be it the advertisers that made commercial broadcasting possible, or the providers indirectly excluded from public funding in a quasi-market system such as that operating in New Zealand. Second, exclusion of final consumers is now possible. New advances in scrambling and encryption technology, and computerised accounting systems, mean that financing broadcasting through private subscription is now feasible—again producing a C mixed good. Subscription TV is a form of club which has already proved effective in private sector provision of popular and highly valued programming such as sports and movies. Can it produce a traditional 'public good'?

If one moves from cell D to cell C, how does the nature of the product change? It is often claimed that private sector defence or security forces provide a rather different nature of good than does a state-funded police force: protection which is both rival and excludable is necessarily of a very different nature from that which is non-rival and non-excludable. One might expect that a more private type of good would emerge, with less universal coverage, less positive externalities, and perhaps fewer complexities such as transactions costs and informational asymmetries. Is the same true of broadcasting?

Certainly the empirical evidence on pay-television suggests that the nature of programming provided would be very different from that currently provided by

public broadcasters such as the BBC (Congdon *et al.* 1992). One point is the lack of universality. Even if one used only the average of the wide range of WTP results, at least 35 per cent of the population would fail to receive BBC programming, which represents a significant welfare loss when one considers that the marginal cost of additional viewers equals zero. These effects would also be exacerbated over time, given the effects on future externalities of consumption and production. Universality used to be so taken for granted as part of the objectives in provision of a definition of a public good that it was often considered unnecessary to even mention it. Now this is no longer the case, and we have to ask ourselves whether exclusion is desirable, given its negative effect on universality. This important moral point relates also to the newly privatised utilities.

Also, one needs to consider what sort of service those who did pay would receive. Would subscription public broadcasting be a perfect substitute for free public broadcasting? Again the consensus is 'probably not'. The experience of existing subscription channels is that viewers are only prepared to pay significant amounts for premium film channels, and not the kind of 'something for everyone' channels which have often typified public broadcasters. Moreover, they tend to have very low programme costs, relying almost exclusively on acquired programming (often purchased from public broadcasters elsewhere). The combination of supply and demand characteristics discussed above means that there does not appear to be the diversity that was once envisaged. Profit-maximising strategies do not seem to lead to the broad mix of entertainment, education and information that one might have expected (Hotelling apart), and so there seems to remain an old-style 'market gap' argument for some other form of provision of broadcasting alongside that provided by the market. Such is the model which has emerged in New Zealand, where some services are treated as belonging to cell D and are tax-funded, while others are advertiser-funded and therefore fit into cell C.

There are earlier precursors of this policy, as other writers have attempted to isolate the extent of the public and private elements of mixed goods, and to price them accordingly. Blaug (1965) and Musgrave (1969) for example, suggest that we separate out the private from the public benefits associated with a mixed good (using the example of education), to calculate a ratio of spillover benefits to private benefits. A ratio between 0 – 1 would be estimated, using cost-benefit analysis, and this could be used to determine public and private pricing. Such an approach is implicit in the New Zealandmodel, where broadcasters essentially have to decide whether an hours' programming should be quango-funded, thus earning a subsidy of *X* dollars to attract a particular type and size of audience; or whether it should be advertiser-funded, thereby attracting a different type of audience and revenue. The decision criterion for the broadcasters, SOE and privately owned alike, is which strategy earns the greater profit.

Another question, of course, is whether programming should be still be provided, if the market appears reluctant or unable to pay for it? That is, given exclusion, should economic profit alone dictate the provision of services to either remote areas, or 'remote' preferences? Interestingly, the Peacock Report has argued that some sort of institutional protection be provided for 'public service'-style programming, alongside its more radical calls for the introduction of subscription, and the quango-style operation we observe now in New Zealand. In part one needs to clearly define what is meant by public service broadcasting. At one level this is easy—it has always meant at the least universal access. However, we must also ask what should it be access to—what is quality broadcasting? What is 'good' about television?

New Zealand has attempted to broach this directly. The 1990 Act which sets out the mandate for the funding quango reveals an interesting shift in the definition of the public aspect of broadcasting. In New Zealand, it is not related to ownership of the service provider, but instead is defined on a programme-by-programme basis, irrespective of who produces or broadcasts it. The state-owned broadcaster is required to act like a commercial enterprise, so much of both private and publicly owned broadcasters' programming is of a C nature. Alongside this, both sectors can provide quango-funded programming of D nature. The 'public good' aspect of broadcasting then appears to be a market gap, externalities and merit goods one, but limited to the individual transaction and not the channel which houses these transactions. The type of transaction is enshrined by law: the quango is to use the licence-fee revenue to promote and develop programmes which reflect 'New Zealand identity and culture'. Such programming would not be expected to be profit-maximising behaviour for a commercially oriented broadcaster. The quango also subsidises provision to geographically remote audiences, thereby honouring the old universality clause.

CONCLUSION

Recent changes in broadcasting policy reflect widespread changes in economic theory pertaining to public goods, and about the role of the state. Increasingly it is argued that state intervention in the funding or provision of public goods cannot be efficient; because government failure undermines theachievable marginal rate of transformation and is worse than the market failure it attempted to remedy, and because the 'preference revelation problem' makes it impossible to know society's marginal rate of substitution for public goods. It is argued that public provision is not the only alternative, as markets can, and in the past did, provide goods with public characteristics. It is even doubted whether public goods exist at all, as the search for the essential element of 'publicness' reduces the concept to such bareness that one must ask what is left. Thus centuries of economists' advice that 'higher authorities' should intervene is undermined.

In broadcasting, as in many other sectors of the public economy, reforms have focused on redefining property rights and introducing competition, even if in a 'quasi-market' form. As in many other sectors, such as health and education, there remain powerful arguments for both continued public funding and for increasing competition at the level of provision. However, these reforms contain many contradictory elements, and it is not clear that markets can adequately replace public provision. There are issues of selective access as compared to universal access to resolve; and also the question of the extent to which the nature of the good is dependent upon the mode in which it is provided. A good which has altered to become more 'private' is not a clear substitute for a former public good.

We do not yet have clear answers to all of these questions, but there is much to be gained by exploring the issues. To this extent, the history of broadcasting economics offers an interesting and insightful metaphor with which to consider the ongoing changes in our treatment of public goods, and of the role of the state. As in all of economics, the differing normative views of the appropriate role of the state can usually be traced back to different positive views about what actually does happen under different institutional structures.

ACKNOWLEDGEMENT

I would like to acknowledge the support and encouragement of my Ph.D. supervisor, Professor A.B.Atkinson.

NOTES

1 Quote attributed (in Briggs 1979) to the editor of the *Manchester Guardian* in the 1930s.
2 For example, New Zealand's 1903 Wireless and Telegraphy Act was one of the first examples of full state monopoly over the yet-to-be-invented broadcasting technology which followed. This was not seen as heavy-handed, rather as appropriate 'hand-in-hand nation building'.
3 Excellent reviews of the public sector experience as a whole include Holtham and Kay (1994).

REFERENCES

Booz, Allen and Hamilton International (UK) Ltd (1987) *Subscription Television: A Study for the Home Office,* London: HMSO.

Blaug, M. (1987) [1965], 'The rate of return on investment in education in Great Britain', in *The Economics of Education and the Education of an Economist,* New York: New York University Press, 3 – 49.

Briggs, A.S.A. (1979) *The History of Broadcasting in the United Kingdom,* vol. IV, Oxford: Oxford University Press.

Coase, R.H. (1988) 'The lighthouse in economies', reprinted in *The Firm, the Market and the Law,* Chicago IL: University of Chicago Press, 187 – 213.

Congdon, T., Sturgess, B., N.E.R.A., Shew, W.B., Graham, A. and Davies, G. (1992) *Paying for Broadcasting: The Handbook,* London and New York: Routledge.

Ehrenberg, A.S.C. and Mills, P. (1990) *Viewers' Willingness to Pay: A Research Report,* London Business School broadcast, London: International Thompson Business Publishing.

Graham, A. (1992) 'The public funding of broadcasting', in T.Congdon *et al. Paying for Broadcasting: The Handbook,* London and New York: Routledge, 167 – 220.

Harcourt, G.C. (1992) 'Markets, madness and a middle way', 2nd Donald Home Address, 17 February 1992, National Centre for Australian Studies, Monash University, Clayton, Victoria. Reprinted in G.C.Harcourt, *Capitalism, Socialism and Post-Keynesianism: Selected Essays of G.C.Harcourt,* Aldershot: Edward Elgar, 9 – 24.

Heilbroner, R. (ed.) (1986) *The Essential Adam Smith,* Oxford: Oxford University Press.

Holtham, G. and Kay, I (1994) 'The assessment: institutions of policy', *Oxford Review of Economic Policy,* 10(3):1 – 16.

Mill, J.S. (1965) *Principles of Political Economy,* Toronto: University of Toronto Press.

Musgrave, R.A. (1969) 'Provision for social goods', in J.Margolis and H.Guitton (eds) *Public Economics: An Analysis of Public Production and Consumption and their Relations to the Private Sectors,* proceedings of a conference held by the International Economic Association, London: Macmillan, 124 – 45.

Noll, R., Peck, M. and McGowan, J. (1973) *Economic Aspects of Television Regulation,* Washington DC: The Brookings Institution.

NOP (1986) *Financing the BBC: A Survey of Public Opinion,* London: HMSO.

Peacock, A. (1979) 'The limitations of public goods theory: the lighthouse revisited', in *The Economic Analysis of Government and Related Themes,* Oxford: Martin Robertson.

Pigou, A.C. (1932) *The Economics of Welfare,* 4th edn, London: Macmillan.

Report of the Committee on Financing the BBC (1986) London: HMSO Cmnd 9824 (Peacock Report).

Samuelson, P. (1954) 'The pure theory of public expenditure', *Review of Economics and Statistics,* (36):387 – 9.

——(1964) *Economics: An Introductory Analysis,* 6th edn, New York: McGraw-Hill.

25
REPETITION, EVOLUTION AND LEARNING IN GAMES
From equilibrium to open-endedness

Giuseppe Ciccarone and Robert Neugeboren

INTRODUCTION

The Prisoner's Dilemma game (PD) models the decision problem of two players, each of whom must decide between two strategies, cooperate *(C)* and defect *(D)*. There are four possible outcomes, and each player evaluates the pay-offs in the following order: $T>R>P>S$, where, for example, S is the pay-off to a player who cooperates when the other defects (Figure 25.1).

The Prisoner's Dilemma poses the problem of cooperation in a simple logical form: given the pay-offs, rational players will choose to defect, and the unique Nash equilibrium *(DD)* is Pareto-inefficient. This analysis has led some to the pessimistic conclusion that cooperation is impossible in a world of rational egoists; that it is, strictly speaking, irrational. Further consideration of the matter, however, seems to point in a more optimistic direction. Examples from biology and elsewhere suggest that strictly selfish behaviour may, over time, take on a socially cooperative form, e.g. the 'selfish gene' gives rise to kinship relations and altruism (Dawkins 1976). While it is true, in other words, that 'two egoists playing the game once will both choose their dominant choice, defection, and each will get less than they both could have got if they had cooperated' (Axelrod 1984:10), there is reason to believe that repeated interaction may produce a different result.

The idea that repetition may resolve the problem of cooperation is not new. At least since Adam Smith, economists have held that market competition tends to 'weed out' inefficient behaviour, so that repeated interactions promote mutually advantageous outcomes in the long run. Darwin was impressed by Smith's description of the market mechanism, and the survival of the fittest is as much a part of economic rhetoric as it is of biological theory. However, the structures and mechanisms involved in biological evolution are better known than their social and economic counterparts,

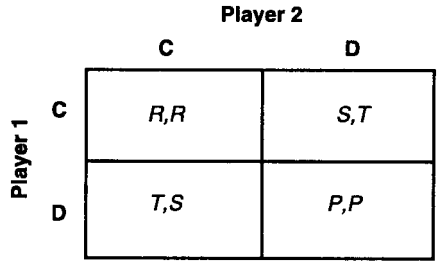

Figure 25.1 The Prisoner's Dilemma

creating a deficiency which has only recently been addressed. The project begun by Von Neumann and Morgenstern (1953)—to 'find the mathematically complete principles which define "rational behaviour" for the participants in a social economy' (*ibid.:* 31)—is today being rehabilitated in evolutionary terms.

Imagine that pairs of strategies from a population are matched at random in games whose outcome determines the number of offspring each leaves to the next generation, so that pay-offs play the role of reproductive rates. Repeated interaction would tend to favour those strategies that do best against the rest, which hence would grow as a share of the population. For example, computer tournaments of the repeated PD have shown that a cooperative strategy can come to dominate the population, because it does better against other cooperators than do defectors against other defectors (Axelrod 1984). In this context, it has been suggested that cooperation may be the outcome of an evolutionary process which 'selects' for equilibria with efficient pay-offs.

There is a tension, however, underlying the attempt to develop evolutionary models within the essentially static terms of equilibrium theory. This arises from the fact that the term 'evolution' is meant to apply to sequences of events where something *new* is coming into being. Evolution cannot be comprehended in an equilibrium world, where 'history' is confined to movements within a stationary system, deprived of a robust sense of novelty and chance. Equilibrium systems lack an evolutionary dimension: they cannot distinguish between past and future, so that processes collapse into a single point in time. Hence what has been missing from the standard approach is an adequate theory of the processes through which equilibrium is achieved, i.e., adaptive behaviour under far-from-equilibrium conditions.[1]

We develop an alternative approach to the problem of cooperation, employing non-linear dynamics to model the growth of a population of rational egoists. In particular, we incorporate the idea that the game is played with fixed resources, so that the population cannot grow without bounds. Our intuition is that if rational players maximise the number ofoffspring they expect to leave to the next generation, then the imposition of an upper limit on the size of the population should affect their expectations and hence their behaviours. By allowing for feedbacks between the behaviour of the players and the growth of

the population over time, we demonstrate that the system can explore qualitatively new dynamics, including the *emergence* of cooperation, i.e. in a population initially dominated by defection, cooperation first gains a foothold and, under well defined conditions, eventually comes to dominate the system.

In our view, evolutionary games are a powerful new paradigm, bringing the analysis of competition and cooperation into the same logical framework. But the intersection of biology and economics may also have introduced an unfortunate confusion. Biological evolutionary theory concerns populations of organisms having the properties of multiplication, heredity and variation (Maynard Smith 1982). Such organisms need not be intelligent, rational or even sentient; what matters is that the carriers of evolution be thought of as self-reproducing programmes or, in game-theoretic terms, strategies. But biological evolution is driven by chance variations which introduce novel strategies into the population, whereas *the engine of social evolution is learning*. Indeed, to the extent that players are able to adapt to inconstant and uncertain environments, selection pressures are effectively forestalled, so that learning from experience itself becomes an advantage.

To understand the processes by which learning takes place, we follow the suggestion of Aumann (1981), Binmore (1987), Rubinstein (1986) and others, and 'seek to model a rational player as a suitably programmed computing machine' (Binmore 1987:181). But our approach differs in the crucial respect that we explicitly admit inductive reasoning (Arthur 1994). Whereas the standard approach employs finite automata to represent immutable repeated game strategies, our machines adapt their behaviours to changes in their environments. Thinking about evolution in an economic setting, in our view, leads away from the traditional ideal of rationality in static, equilibrium terms and towards a notion of a learning process that adapts to its environment and hence changes over time (Rapoport 1966:210).

In a simple duopoly model, for example, we can imagine that firms are represented by finite machines playing a repeated game with incomplete information. Each player is part of the environment for the other, which is thus not given at the outset but is found out in the process of playing the game, whereby one reacts to what one learns and by these reactions modifies the environment (Rapoport 1966). Through repeated interaction, each player forms expectations of the other's output, or behaviour, which is a choice of moves in the game, revising these expectations in light of correct or disappointed predictions. Given this set-up many outcomes are possible, and the dynamics of the game will be path-dependent, since small differences in initial conditions may lead to the cooperative rather than thenon-cooperative solution, as both Rapoport and Chammah (1965) and Smale (1980) have shown for special cases.

Nonetheless, we hope to point to a rather general phenomenon: cooperation entails a qualitative change in the internal organisation of the system which

cannot be reduced to a choice of strategies. While individuals are the units of competition, it is the group that 'learns' to cooperate (Samuelson 1993).

The paper is organised as follows. In the next section, we consider the main experimental results on cooperation in repeated games (Axelrod 1984) and propose an interpretation we believe has been neglected by the literature. The third section discusses evolutionary game theory. The fourth section introduces some formalism based on the analysis of finite automata. The fifth section adapts this formalism to account for learning. The sixth section presents a model of population growth under a resource constraint and defines conditions on the pay-off matrix for which automata 'learn to cooperate'. Finally we offer some interpretation and concluding remarks.

COOPERATION IN REPEATED GAMES

Axelrod (1984) describes a pair of computer experiments where opponents played a repeated PD[2] in a series of round-robin tournaments.[3] Contestants submitted a programme which selects a move (*C* or *D*) on each play of the game, and each entry was matched against every other, itself and a control, RANDOM. The tournaments were won by the entry of Anatol Rapoport, TIT FOR TAT *(TFT),* which cooperates on the first move and then does whatever its opponent did on the previous move. 'This was the simplest of all submitted programs and it turned out to be the best!' (*ibid.:* 31). But if defection is strictly dominant in the one-shot game, why should repeating the game lead to a different outcome? And if a finitely repeated PD leads to defection on the last and hence, by backward-induction, on each move of the game, how can cooperation emerge 'if the players will interact an indefinite number of times' (*ibid.:* 10)?

Axelrod offers an explanation for the success of *TFT* in terms of the necessary and sufficient conditions for a strategy to be collectively stable (*ibid.:* 57 – 9, 207 – 11). *TFT* is collectively stable if no other strategy can invade it.[4] This definition relies on the calculation of the sum of an infinite series of discounted pay-offs. The *discount parameter, δ,* is the weight 'of the next move relative to the current move…the degree to which the pay-off of each move is discounted relative to the previous move' (*ibid.:* 13).[5] In essence, if the discount parameter is sufficiently high, then the best a player can do against an opponent using *TFT* is to adopt the strategy herself, with the result that both will cooperate on every move of the game. In other words, if $\delta > \delta^*$, *TFT* 'is in Nash equilibrium with itself' (*ibid.:* 217, n1).

In repeated games, each player must consider the reactions of her opponent(s) in choosing a strategy, so that if the 'shadow of the future' looms large, the long-run benefit of cooperation may outweigh the short-run temptation to defect. In this context, what *TFT has* going for it is a high degree of reciprocity, making each move contingent on the opponent's prior move. Where it is possible for players to punish each other for defection in the present, it may be possible to

promote cooperation in the future. The fear of retaliation may thus lead to outcomes that otherwise would not occur' (Fudenberg and Maskin 1986:533). Indeed, the Folk theorem establishes that any feasible individually rational pay-off may be supported as an equilibrium of the repeated game, if players are 'sufficiently patient' *(ibid.).* Thus while *TFT is* in equilibrium with itself, so too is always defect *(ALLD)* and a host of other strategies.

It should be clear, therefore, that *TFT* is collectively stable is not sufficient to explain the *emergence* of cooperation in the repeated game.[6] To clinch the point, consider the choice between two strategies: *TFT* and *ALLD.* Restricting attention to these two strategies is easy to justify: they encapsulate cooperative and non-cooperative behaviour in the repeated game and are sufficient to illustrate the nature of Axelrod's analysis without loss of generality. The repeated game pay-off to someone using *ALLD* when playing with someone using *TFT, π(ALLD \TFT),* is the sum of *T* for the first move, δP for the second, $\delta^2 P$ for the third, and so on (Axelrod 1984: 13 – 14).

Consider what happens to the repeated game pay-offs as the discount parameter approaches unity (i.e. $\delta \rightarrow 1$). Where $\delta=0$, the uture is of no importance, and the analysis reduces to the one-shot case, where defection is strictly dominant. However, as δ increases beyond its critical value, δ^*,[7] *the pay-off structure changes*. That is, where $\delta<\delta^*$, the following inequalities hold: π *(ALLD\TFT)>π(TFT\TFT)>π(ALLDIALLD)>π(TFT\ALLD).* And where $\delta>\delta^*$, the following inequalities hold: *π(TFT\TFT)> π(ALLD\TFT),>π(ALLD\ALLD)>π (TFT\ALLD).* Where the discount parameter is below its critical value, δ^*, the pay-off to *ALLD* is greater than that to *TFT,* no matter what the opponent does. If $\delta>\delta^*$, then *π(ALLD/ALLD)>π(TFT\ALLD)* but *π(TFT\TFT)>π(ALLD\TFT),* so that when the discount parameter is above its critical value *there is no dominant strategy*.[8] In this case, the game is not a PD, but a related one-shot game, the Assurance Game (AG),[9] where two equilibria (one inefficient and one efficient) exist, giving rise to a coordination problem.[10] Then a player will choose *TFT* if and only if, first, $\delta>\delta^*$, and, second, she expects the other to do likewise.

Thus in posing the problematic, 'the evolution of cooperation', Axelrod suggests a historical dimension which is in fact absent from his experimental set-up; once the discount parameter is known, each player announces her strategy and the outcome can immediately be calculated. His theoreticalanalysis likewise remains bound to the conceptual scaffolding of an essentially static, equilibrium world. Indeed, the condition under which *TFT* is collectively stable does not explain how it is that cooperation 'evolves' through time; rather it identifies, *as proved by the Folk theorem,* a large set of outcomes from which, once achieved, neither player has an incentive to deviate. Efficient and inefficient equilibria are possible, and how players come to expect cooperation remains an open question.

EVOLUTIONARY GAME THEORY

The success of *TFT* in computer experiments has led some authors to use it as 'a synonym for a self-enforcing, cooperative agreement' (Binmore 1991: 433). But experimental results are very sensitive to the way in which the model is specified, since the winner of any actual contest will depend on the set of strategies present and on the length of play. For example, Young and Foster (1991) simulated Axelrod's tournament, increasing the number of iterations from a few hundred to 100,000. They show that while *TFT does* best in the short-run, any other initially disadvantaged strategy may come to dominate the system in the long-run, depending on what strategies are present. On experimental grounds, therefore, 'one cannot conclude very much about the long-run viability of any one strategy without making a detailed study of the potential function associated with the whole system' (*ibid.:* 154 – 5).

Theoretical analyses, on the other hand, run into the infinitely large set of Folk theorem equilibria. In this context, it has been suggested that cooperation may be the outcome of an evolutionary process that selects against inefficient behaviour, following the 'widespread intuition that often the most likely equilibria are those whose payoffs are efficient' (Fudenberg and Maskin 1990:274). Yet deterministic selection dynamics cannot discriminate between strict equilibria, so that the entire set of Folk theorem equilibria remain relevant. In other words, nearly anything can happen. But if the system is perturbed, then under well defined conditions the set of repeated game equilibria may 'settle down' to include only efficient elements. Binmore and Samuelson (1992) distinguish two approaches: first, *internal* trembles, which correspond to mistakes at the level of the 'thinking process of the players' (*ibid.:* 284); second, *external* trembles, which correspond to indeterminate (i.e. noisy) environments.

For example, Fudenberg and Maskin (1990) demonstrate that evolutionary stability (ES) implies efficiency, if players make mistakes with a small but positive probability. Hence efficient equilibria may arise as the result of a weeding-out process, if 'the penalty for trying to initiate efficient play against an inefficient opponent is not large' (*ibid.:* 275). Fudenberg and Harris (1992) model external trembles as exogenous shocks to the aggregate growth rates(e.g. 'weather'), introducing 'the stochastic elements at the level of the growth rates of individual populations' (*ibid.:* 421). (We take this as the point of departure for the model presented in the penultimate section below.)

By adding a stochastic element to the selection dynamics, it is possible to define a process such that as the error term goes to zero, the solution converges to the set of efficient equilibria. This route has been taken by many economic theorists who seem determined to show that 'utility maximising behaviour on the part of each individual *necessarily* leads to cooperation' (Aumann and Sorin 1989:6). Yet to maintain that *only* efficient strategies survive repeated plays of the game puts too much of the burden on notions such as noise, mistakes, trembles, etc. Such an approach only obscures the fundamental tension between

rationality and efficiency posed by the one-shot game, *in which those notions have no logical analogues*. Evolutionary games incorporating stochastic elements proliferate, but as in the older literature on trembling-hand perfect equilibria (Selten 1975) little explanatory power is gained if cooperation is founded on non-rational, essentially chance events. For once the noise and mistakes are eliminated, so too is the basis for cooperation (Sen 1985:344).

In our view, the idea that 'evolution implies efficiency', which motivates much of this research, has simply transferred its faith from the beneficence of the invisible hand to the beneficence of natural selection. But evolution requires that players adapt to inconstant and uncertain environments, where learning is the key to success. Hence what is missing from the theory of evolutionary games is an adequate conceptualisation of the way players acquire and process information over the course of repeated interaction, i.e. the entire process that intervenes between the receipt of a decision stimulus and the ultimate behaviour, including the manner in which the decision-maker forms expectations and adapts to changes in the environment. In this context, we are led to model players as suitably programmed machines.

RATIONAL PLAYERS AS FINITE AUTOMATA

In a repeated PD with a finite number of rounds, the only Nash equilibrium has each player defecting in every round. In this context, it has been proposed that 'uncooperative behaviour is the result of "unbounded rationality", i.e. the assumed availability of unlimited reasoning and computational resources to the players' (Papadimitrou 1992:122). Hence the traditional commitment to utility maximisation has given way to considerations of the finite computational abilities of economic agents.

Mathematicians interested in modelling reasoning systems employ the idea of a Turing machine. Turing had been interested in the question: 'is there some general mechanical procedure which could, *in principle,* solve all the problems of mathematics (belonging to a suitably well-defined class) oneafter the other?' (Penrose 1989:34). To put the question in an unambiguous form, he devised a machine possessing a finite number of internal states fed instructions on an input tape comprised only of 0's and 1's. Any possible design of such a finite machine is specifiable by a natural number, referred to as its Gödel number. Turing proved that such a machine is in principle capable of any formal calculation that a human mathematician can accomplish in a finite number of steps, i.e. one can construct a 'universal' Turing machine which, if offered data *d,* and the Gödel number of any other machine *X,* produces the same output as *X* would given data *d*. In game-theoretic contexts, the kind of recursive reasoning involving chains of I-think-that-you-think-that-I-think, etc., can be precisely formulated (a stopping problem), and pairs of machines can be analysed in straightforward terms.

In what follows, we employ a simple automaton, introduced to game theorists by Rubinstein (1986). A Moore machine possesses a finite number of internal states, one of which is the initial state, an output function and a transition function. For the analysis of the PD (or any 2×2 game), we need to consider a machine with at most two internal states and a simple output function. In one state the machine always outputs *C* and in the other always *D*. The transition function describes how the machine is caused to change states by its interaction with the outside world. Machines meet in pairwise interactions and play at each round either *C* or *D*. Depending upon the outcome of that round, they 'decide' what to play on the next round, and so on.

Formally, a Moore machine (M_i) is a four-tuple $(Q_i, q_i^0, \lambda_i, \mu_i)$. Q_i is the finite set of the internal states of M_i, and $q_i^0 \in Q_i$ is the initial state of M_i. $\lambda_i: Q_i \rightarrow$ *i,* so that $s_i^t = \lambda_i(q_i^t)$ is a move in PD that M_i chooses when it is at state q_i^t; λ_i is called the output function. $\mu_i: Q_i \times \ _i \rightarrow Q_i\ (j \neq i)$; μ_i is called the transition function. At time t=1, M_i starts at state $q_i^1 = q_i^0$ and $\lambda_i(q_i^1) = s_i^1 \in \Phi_i$.At time *t,* $\lambda_i(q_i^t) = s_i^t \in \Phi_i$ and $\mu_i(q_i^t, s_j^t) = q^{t+1}_i$

Some examples are given in Figure 25.2.[11]

LEARNING AUTOMATA

In the standard analysis, players choose machines once and for all, so that 'time has almost an artificial role' (Smale 1976:289). Players know the pay-off matrix and outcomes can immediately be calculated. In what follows, we are instead concerned with the way machines adjust their behaviours over time, where the pay-off matrix is unknown. Machines are described by probabilistic transition functions, and behaviour is conditioned on the (possibly infinite) history of their interactions. Multiple solutions exist and, unlike the Folk theorem, convergence is obtained in real time.

Rapoport and Chammah (1965) and Rapoport (1966) pioneered this inductive approach to finite automata in repeated games. They consider asimple learning process based on the assumption that $P=-R$ and $S=-T$: automata start with a random choice of strategies and repeat it if the pay-off is positive, change it if the pay-off is negative.

For i=(1, 2), define as x_i the probability that player i will choose *C* after *CC*, as w_i the probability that she will choose *C* after *DD,* as y_i the probability that she will choose *C* after *CD,* as z_i the probability that she will choose *C* after *DC*. Call these the *conditional probabilities*. By using Moore machines, it is easy to verify that if x_i=1, w_i=1, y_i=0, z_i=0, the solution converges (Smale 1980) to [*(CC), (RR)*] in at most three periods.

This machine is:

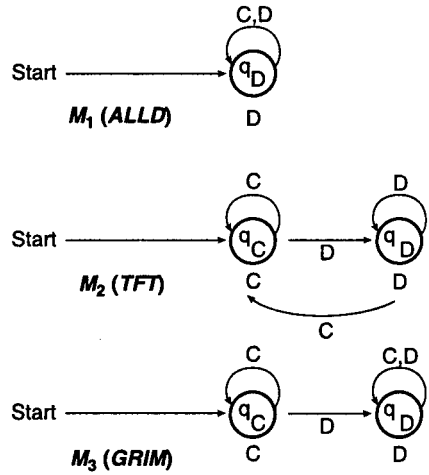

Figure 25.2 The Moore machine: some examples

$$M_i = (Q_i, q_i^0, \lambda_i, \mu_i), \text{ for } q_i^0 \in Q_i \qquad \lambda_i(q_i^t) = s_i^t \in \Phi_i$$
$$\mu_i(q_i^t, s_j^t, x_i = 1, w_i = 1, y_i = 0, z_i = 0) = q_i^{t+1}$$

$$Q_i = (q_{iC}, q_{iD}) \qquad q_i^0 = q_i^0 \qquad \lambda_i(q_{is}^t) = s_i^t$$
$$\mu_i(q_{it}, s_j^t, \cdot, \cdot, \cdot, \cdot) = \begin{cases} \mu_i(q_{iC}^t, \cdot, \cdot, \cdot, \cdot, \cdot) = (q_{is}^{t+1}) \\ \mu_i(q_{iD}^t, s_j^t) = C(\cdot, \cdot, \cdot, \cdot, \cdot) = (q_{iD}^{t+1}) \text{ for } s = (C, D) \\ \mu_i\ (q_{iD}^t, s_j^t = D, \cdot, \cdot, \cdot, \cdot) = (q_{iC}^{t+1}) \end{cases}$$

The transition diagram is provided in Figure 25.3.

Rapoport and Chammah (1965) also consider a more complicated learning process. In particular, they assume that x_i is adjusted at a rate proportional to the gradient of the expected gain:

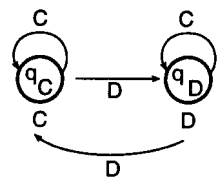

Figure *25.3* The Position

$$\dot{x}_i = \frac{dx_i}{dt} = k_i \frac{\partial G_i}{\partial x_i}$$

Now w_i=1, y_i=0, z_i=0, and M_i will choose C after CC with a variable probability $0 \leq x_i^t \leq 1$

This machine can then be written as:

$$M_i = (Q_i, q_i^0, \lambda_i, \mu_i) \quad \text{for } q_i^0 \in Q_i \quad \lambda_i(q_i^t) = s_i^t \in \Phi_i$$

$$\mu_i(q_i^t, s_j^t, x_i^t, w_i = 1, y_i = 0, z_i = 0) = q_i^{t+1}$$

After computing the steady-state Markov equations and solving for the steady-state distribution, the expected gains to the two players and their first derivatives are calculated. Call x^* the unique value belonging to the unit interval (for T<3R) for which

$$\frac{\partial G_i}{\partial x_i} = 0$$

For $x_i > x^*$,

$$\dot{x}_i = k_i \frac{\partial G_i}{\partial x_i} > 0$$

the adjustment process will lead to x_i=1, and convergence to the cooperative solution *[(CC), (RR)]* occurs at most in three periods. For $x_i < x^*$,

$$\dot{x}_i = k_i \frac{\partial G_i}{\partial x_i} < 0$$

the adjustment process will lead to x_i=0, and the process will continuously oscillate between s_t=*(CC)* and s_{t+1}=(*DD*).

This machine is:

$$M_i = (Q_i, q_i^0, \lambda_i, \mu_i) \quad \text{for } q_i^0 \in Q_i \quad \lambda_i(q_i^t) = s_i^t \in \Phi_i$$

$$\mu_i(q_i^t, s_j^t, x_i = 1, w_i = 1, y_i = 0, z_i = 0) = q_i^{t+1}$$

$$Q_i = (q_{iC}, q_{iD}) \qquad q_i^0 = q_i^0 \qquad \lambda_i(q_{is}^t) = s_i^t$$

$$\mu_i(q_i^t, s_j^t, \cdot, \cdot, \cdot, \cdot) = \begin{cases} \mu_i(q_{iC}^t, \cdot, \cdot, \cdot, \cdot) = (q_{iD}^{t+1}) \\ \mu_i(q_{iD}^t, s_j^t = C, \cdot, \cdot, \cdot) = (q_{iD}^{t+1}) \quad \mu_i(q_{iD}^t, s_j^t = D, \cdot, \cdot, \cdot) = (q_{iC}^{t+1}) \end{cases}$$

The transition diagram is provided in Figure 25.4.

The cycle arises if w_i=1. If instead w_i=0, the solution converges to the non-cooperative solution *[(DD), (PP)]*. In general, where two states are considered, convergence to either the cooperative solution or to the non-cooperative solution

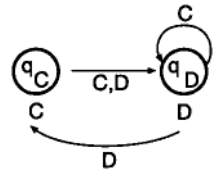

Figure 25.4 The cycleton

depends on the initial values of the conditional probabilities and the way they are updated in the course of pairwise interaction. This result shows that the problem of multiple equilibria may be resolved in real time, without imposing arbitrary bounds on the set of available strategies. The initial conditions of the system and its law of motion make more than one evolution possible, however, and the dynamics do not necessarily converge to stationary modes of behaviour (and hence solutions). Therefore we now turn our attention to the question whether there exists a general class of processes that guarantee convergence.

It can be easily verified that from any initial conditions convergence of the solution to *[(DD), (PP)]* requires: $x=w=y=z=0$. If $x=w=1$ and $y=z=0$, then the solution will converge to *[(CC), (RR)]*. Hence if two probabilities (x and w) are treated as variables, the machine can converge to either the cooperative or the non-cooperative solutions.

The machine is then:

$$M_i^* = (Q_i, q_i^0, \lambda_i, \mu_i), \quad \text{for } q_i^0 \in Q_i$$
$$Q_i = (q_{iC}, q_{iD}) \qquad q_i^0 = q_{iD}$$

$$\lambda_i(q_i^t) = s_i^t \in \Phi_i \qquad \mu_i(q_i^t, s_j^t, x_i^t, w_i^t, y_i = 0, z_i = 0) = q_i^{t+1}$$

The corresponding transition diagram is provided in Figure 25.5.

Each machine, i, may update the conditional probabilities differently, but in general the dynamics of x and w can be described by the following one-dimensional iterative maps:

$$x^{t+1} = f(x^t) \quad 0 \le x \le 1$$
$$w^{t+1} = h(w^t) \quad 0 \le w \le 1$$

where $f(.)$ and $h(.)$ are assumed to be continuous and continuously differentiate, $x^*=0$, $x^*=1$, $w^*=0$, $w^*=1$ are stable fixed points and, if there exists another pair of vectors $x^*=(x_1,\ldots, x_M)$ and $w^*=(w_1,\ldots, w_Q)$ of fixed points, then they are unstable.[12]

$$x^{t+1} = f(x^t) \quad 0 \leq x \leq 1$$

Figure 25.5 The learning automaton

Of course, many functions $f(.)$ and $h(.)$ can be specified (two examples are depicted in Figure 25.6). For our purposes, we need only find the general form of the curve representing the unknown functions (Poincaré 1909).

By using Rapoport and Chammah's (1965) conditional probabilities, different values for some parameters of the transition function make convergence possible, for example by moving away from cycles. On the basis of this result, we have constructed a class of automata able to bring about convergence to either the cooperative or the non-cooperative solution. These automata will be employed in the next section to characterise the emergence of cooperation in an environment with a resource constraint. The population is comprised of heterogeneous machines characterised by different functions sharing the same general form. For any such functions, $x\ (w)$ will converge to 1 if the initial x^1 (w^1) is greater than the unstable fixed point; to zero in the opposite case. The solution is sensitively dependent on the initial conditions, since small differences may lead to the cooperative rather than the non-cooperative solution, as both Rapoport and Chammah (1965) and Smale (1980) have shown for special cases, and as we have already stressed in the introduction.

LEARNING TO COOPERATE (UNDER FIXED RESOURCES)

In what follows, machines adjust their conditional probabilities and acquire behaviours that are passed on to later generations. Each machine's transition function describes how it is caused to change states by its interaction with the outside world. On the one hand, machines meet in pairwise interactions and play at each round either *C* or *D*. Depending upon the outcome of that round, they 'decide' what to play on the next round, and so on. The history of this interaction will converge to one of two outcomes, *CC* or *DD,* within a single generation. On the other hand, this history is embedded in a larger one. Each generation, through its utilisation of environmental resources, determines the initial conditions for the next. Hence, machines receive a signal from the environment about the overall dynamics of the system, so that the transition function reflects the history of machine/machine and machines/environment interactions.

Consider the growth of a population of learning machines in an environment characterised by a fixed carrying capacity *(K)*. The population is initially dominated by defectors. In each generation $(g=1,..., G)$, individuals engage in a series of PDs lasting t rounds $(t=0,\ldots, T)$, and the outcome determines the number of offspring left to the next generation.[13] If resources are unlimited and generations do not overlap, the number of individuals (N_g) grows at the constant

rate of reproduction *(r)* corresponding to the outcome of the repeated PD. If resources are instead limited, and individuals utilise food, etc. to live and reproduce, the system can only maintain a fixed number of individuals, i.e. the population's size is upwardly bounded by the system's carrying capacity.

The discrete-time formulation of this process is given by the logistic difference equation:

$$N_{g+1} = N_g[1 + r(1 - N_g/K)] \tag{1}$$

As the ratio of the number of individuals in the environment to its fixed carrying capacity increases $(N_g/K \rightarrow 1)$, the size of the population approaches its steady-state value $(N^*=K)$ only for $0 \leq r \leq 2$ (May 1976).

Defining

$$A = \frac{r/K}{1+r} N$$

equation (1) can be brought into canonical form:

$$A_{g+1} = (1 + r)A_g(1 - A_g) \tag{2}$$

Call A the *population density*. The basic fixed point of Equation (2) is represented by A^*. If $0 \leq r \leq 1$, the population will smoothly increase until reaching the stable fixed point; if $1 < r \leq 2$, the population will exhibit cobweb time paths whose oscillations converge to the stable fixed point; if $r > 2$, the fixed point is unstable and all sorts of dynamics may be generated, including cycles and possibly chaos. The rate of growth of the population hence depends upon the rate of reproduction, which is determined by the average pay-off (P or R) of each repeated game, and hence by the behaviour of players.

Logistic growth implies that when the size of the population relative to the carrying capacity is small, the number of offspring each individual will have is a positive function of its own reproductive success and the reproductive success of the other(s), since more individuals will be around for future interactions. Once the density of the population reaches a certain critical value, however, competition for scarce resources takes over and the dynamics become more complex. With the onset of complex dynamics the system can explore qualitatively new paths, and the number of offspring left to the next generation will be hard to predict. Nonetheless, while defection will predominate in early generations, mutually rewarding patterns of cooperation can become established and drive the system toward the efficient steady state.

To illustrate this in a simple way, we assume that at $t=1$, $g=1$, machines receive the input: $x^*=w^*=A^*$. At round t of generation g, the conditional probabilities are defined as x_g^t and w_g^t. At round $t=1$ of generation g, machines receive the input: $x_g^1 = w_g^1 = A_g$.. Together, these assumptions encapsulate the ability of machines to adapt to changes in population density.

In each generation a series of repeated PD is played, and all outcomes—given x^*, w^*, x_g^1 and w_g^1 will converge to either *DD* or *CC*. The associated average pay-off (*P* or *R*) determines the rate of reproduction for generation *g* and, by equation (2), A_{g+1}. At round $t=1$ of generation $g+1$. the machines receive the input $x_{g+1}^1 = w_{g+1}^1 = A_{g+1}$. Given x^*, w^*, x_{g+1}^1 and w_{g+1}^1 the outcome will again converge to either *CC* or *DD*. This determines the rate of reproduction of generation $g+1$ and A_{g+2}, and so on. The conditions for the population's smooth convergence to the steady-state value can be given in terms of the pay-offs matrix, and we are able to distinguish two regimes around a knife-edge 'separating a trend toward maximum cooperation from a trend toward minimum cooperation' (Rapoport 1966:156).

The predominance of defection: $0 \leq P \leq 1$

Machines receive the input $x^*=w^*=A^*$ and start in $q_i^0 = q_D$; hence, the outcome of the first round of play is *DD*. They then receive the input $x_g^1 = w_g^1 = A_g < A^*$. Given this input, the interaction continues until the solution converges to *[(DD), (PP)],* so that the first generation's rate of reproduction is *P*. A_{g+1} is then determined by Equation (2). As A_g approaches A^*, $x_g^1 < x^*$ and $w_g^1 < w^*$ for all $g < \infty$ Hence, given the initial values of the conditional probabilities in each generation, both variables will tend to zero. In other words, each interaction may produce some cooperation early on, but it will eventually dampen down to *DD*. The rate of reproduction will always be equal to *P,* and the population will smoothly evolve toward N^*.

The emergence of cooperation: $P>1$

Machines receive the input $x^*=w^*=A^*$ and start in ; hence $q_i^0 = q_D$ the outcome of the first round of play is *DD*. They then receive the input $x_g^1 = w_g^1 = A_g < A^*$. Given this input, the interaction continues until the solution converges to *[(DD), (PP)],* and the first generation's rate of reproduction is *P*. A_{g+1} is then determined by equation (2). As A_g approaches A^*, x_g^1 approaches x^* and w_g^d approaches w^*. However, at some point along the path it will happen that $x_g^1 > x^*$ and $w_g^1 > w^*$, so that the conditional probabilities will tend toward unity and the solution will converge to *[(CC), (RR)].* The logistic equation will be tuned by *R* with a fixed point $A^{**}>A^*$. Note that the precise evolution of cooperation is difficult to specify in general terms, since it will also depend on the value of *R* (the rate of reproduction of the system for generations converging to the cooperative solution).[14]

This result is obtained by linking the population dynamics to the initial values of the conditional probabilities in each generation and by imposing a 'tight' relationship among x^*, w^* and A^*. Thus the complete evolutionary path of the system is the product of two processes: a (fast-moving) adjustment process

whereby conditional probabilities change over the course of many pairwise interactions; and a (slow-moving) adjustment process whereby the size of the population approaches the carrying capacity of the environment. The emergence of cooperation can be graphically depicted (an example is provided in Figure 25.6).

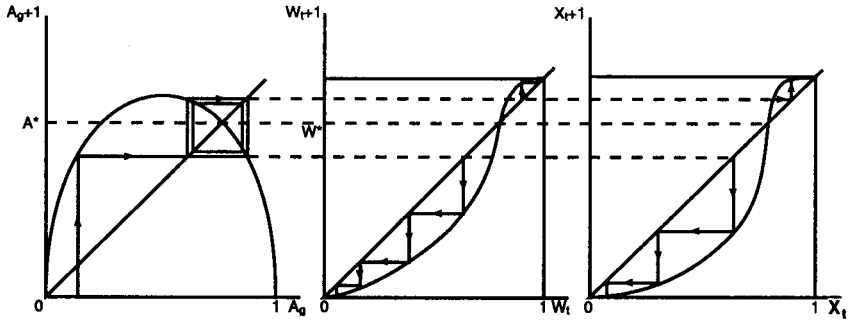

Figure 25.6 The emergence of cooperation

CONCLUDING REMARKS

Equilibrium theory is the deepest, best worked out part of economics. Yet the formalisation and progressive clarification of fundamental concepts that game theory has provided has, in turned, raised questions calling the usefulness of equilibrium theory itself into question. Indeed, Von Neumann and Morgenstern (1953:44 – 5) saw their work as preliminary to a more adequate, dynamic theory, and a critical mass of contemporary theorists is forming around this project today. Our own effort follows from the literature on the repeated Prisoner's Dilemma, the analysis of which has gone beyond traditional equilibrium models and introduced evolutionary concepts and concerns. In particular, achieving cooperation involves motions that carry the system away from equilibria that are suboptimal, and these can only be understood within the conceptual scaffolding of an essentially dynamic model. But rather than interpret these motions as accidental or chance events (e.g. mistakes or mutations), we see cooperation as the outcome of an intentional, cognitive process.

Introducing Moore machines that adjust their conditional probabilities is a step towards understanding the cognitive processes (e.g. learning) involved in achieving cooperation in a repeated PD. In our model, cooperation is 'triggered' by endogenous fluctuations generated by the deterministic dynamics of the growth of a population under a resource constraint. This approach allows us to model feedbacks between the behaviour of the individuals and the size of the niche over which they compete. The multiple effects of many pairwise

interactions determine the rate of population growth, yet limited resources constrain the elementary behaviour of the individuals.

We have specified conditions on the pay-off matrix for which the one-shot game occurs as a special case: if only one generation and one round occur, then machines play *D* with certainty. Moreover, if $0 \leq P \leq 1$, any series of outcomes will dampen down to *DD*. Both cases are consistent with the existence of a dominant strategy, and hence with the 'hard core' of game-theoretic reasoning. However, if $P > 1$, the outcome will converge to *DD* in early generations but will eventually converge to *CC* in some later generation. One interpretation of this result is that if a generation's pattern of behaviour allows the system to smoothly reach the steady state (i.e. without oscillations), the next generation will reproduce the previous behaviour, and so on. If the resource constraint is tighter, however, then the system will be driven to explore new patterns, including cooperation.

Deploying automata in evolutionary games offers a mathematically rigorous way to model the open-ended processes which characterise adaptation in natural and artificial systems (Holland 1975). In this context, it should be noted that Von Neumann devoted much of his work after the *Theory of Games* to the logical description of a self-reproducing automaton(Von Neumann 1987:391 – 552). 'This was primarily an abstract model that could be precisely formulated in mathematical terms, but it was based on a number of presuppositions that have also been recognised as essential to the self-organisation of living systems' (Eigen and Winkler 1983:189). Von Neumann imagined a machine floating in a lake in which also were floating all the parts necessary for its reconstruction. It would gather and assemble these parts according to its own blueprint, and it would copy the blueprint and insert it into the new entity, which would then be able to repeat the procedure. The entire programme would thus be capable of self-reproduction, and '[s]elective changes in the programme could produce constant improvement in the machine and an extension of its capabilities' *(ibid.)*.[15]

In our view, traditional thinking about the 'principles that define "rational behaviour" in a social economy' has been severely limited by the static nature of equilibrium theory. Evolutionary theory, on the other hand, implies that the system under consideration is in principle open-ended; a closed system cannot evolve.[16] Evolutionary systems are 'open-systems-in-environment' (Luhmann 1982), both incomplete and self-referential, simultaneously open and closed: 'open' because they are capable of adapting to a uncertain and inconstant environment, and 'closed' because this adaptation can only take place according to their own internal patterns and decision algorithms. But even 'programmed' machines may exhibit surprising behaviour, if their interactions are sufficiently complex.

In our model, learning takes place as a matter of course (for $P > 1$), but in interpreting this result we should not forget the possibility of not learning. That the condition calls for cooperation and whether in fact it occurs are separate matters. More generally, because players' interactions change their information

sets and hence their learning processes, it is an open question whether the dynamics will converge to a specific steady state. If the evolution of the game does not generate signals which enable players to believe that a path-independent state will be reached, then expectations about others' path-expectations will become important. If, for example, there are many stationary cycles or 'chaotic' time paths, each player will need to try to understand what the other player believes is going to happen, and the formalisation of their learning process may become a difficult task to perform (Grandmont and Malgrange 1986).

Hence a theory of the learning processes involved may not be an easy matter; in some cases, it would be simply impossible to say what theory players should hold and how they should update it. This difficulty might well make cooperation impossible—unless, that is, organisational structures emerge which solve the problem, for example by distributing information among the players so as to influence their 'reputations', or by promoting conventions or rules of thumb. But much more work needs to be done to understand the emergent properties of even simple economic systems.

A paradox cannot be solved; rather it indicates there is something wrongwith the formulation of the problem. Our 'resolution' to the PD is a dynamic model which has the one-shot version as a special case. We have shown that the emergence of cooperation can be explained by incorporating the idea that the game is played with fixed resources. The next step is to investigate what happens when the carrying capacity of the system is variable. In that case, the evolution of cooperation would depend on introducing new ways of exploiting existing resources or of finding new ones.

ACKNOWLEDGEMENTS

We wish to thank A.Rapoport, G.Dymski, R.Gibbons, E.Maskin and A. Rubinstein for valuable discussions and comments on earlier versions of this paper. The usual disclaimer applies.

NOTES

1 As Von Neumann and Morgenstern (1953:44 – 5) warned: 'We repeat most emphatically that our theory is thoroughly static. A dynamic theory would be more complete and therefore preferable…. A static theory deals with equilibria. The essential characteristic of equilibrium is that it has no tendency to change, i.e., that it is not conducive to dynamic development. [But] for the real dynamics that investigate the precise motions, usually far away from equilibria, a much deeper knowledge of these dynamic phenomena is required'.

2 Formally, a Prisoner's Dilemma is defined as PD=($_1$, $_2$, u_1, u_2). The strategy set for player i is: $_i$=*(C, D),* for *i=(1, 2),* and the pay-off function is the map u_i: $_1 x$ $_2 \rightarrow$• ^{2}. In a repeated PD, at each time t player i chooses $s_i^t \epsilon \Phi_i$ and observes the opponent's strategy. is the convex closure of the points *(TS), (RR), (ST), (PP)* and

the pay-off vector at time t is: $u^t = (u_1^t, u_2^t) \epsilon \Gamma$. An outcome at time t is the pair: $s_t = (s_1^t, s_2^t)$.. A solution is an outcome and a pay-off vector, where the pay-off vector is stationary for the dynamics defined by s_t.

3 In the first round of the tournament the game lasted exactly 200 moves, while in the second round the length of the game varied probabilistically with an average of 151 moves.

4 Strategy *X* is said to invade strategy *Y*, if *X* receives a higher pay-off when playing with someone using *Y* than does *Y* when playing with someone using *Y*. This differs from the notion of an evolutionarily stable strategy (ESS).

5 An alternative interpretation of the discount parameter is the likelihood of the game continuing for another move, so that the expected number of moves per game is $\frac{1}{1-\delta}$.

6 Of course, to have shown that *TFT* is a collectively stable strategy is not equivalent to a prediction of a winner in the computer tournaments. That *TFT* will not necessarily win is proved by a trivial counter-example: if only two entries, *TFT* and *ALLD,* had been submitted to round one of the tournament, *ALLD* would have won. In fact, *TFT* would loose in *any* tournament with only two entries. In practice, *TFT* was placed second in a preliminary tournament (Axelrod 1984: 32).

7
$$\delta^* = \frac{(T - R)}{(T - P)}$$

It should be noted that δ^* is the same critical value as Axelrod's (ch. 1 , n5).

8 If $\delta=\delta^*$, then: $\pi(TFT\backslash TFT)=\pi(D\backslash TFT)>\pi(D\backslash D)>\pi(TFT\backslash D)$. It is impossible to characterise this case in the static terms of an equilibrium model, since it exists only as a 'point of flux' between a Prisoner's Dilemma and an Assurance Game.

9 An Assurance Game is defined as one in which $R>T>P>S$ (Sen 1967).

10 In a private communication to the authors, Anatol Rapoport suggested: 'the equilibrium *TFT/TFT* in the Assurance game Pareto-dominates the equilibrium *ALLD/ALLD*. By invoking Harsanyi's method of deciding among multiple equilibria, one can immediately designate *TFT/TFT* as the only rational one'. This approach avoids some of the difficulties associated with multiple equilibria (i.e. coordination problems and mutual expectations), however formalising the common knowledge required raises its own difficulties. See, for example, Bicchieri (1987) and Binmore (1987).

11 The transition diagrams follow Rubinstein's (1986: section 2.3) conventions.

12 Formally, to specify the required functional forms for $f(.)$ and $h(.)$, the following conditions must hold:

$$f(0) = 0 \qquad h(0) = 0 \qquad |f'(0)| \leq 1 \qquad |h'(0)| \leq 1$$

$$f(1) = 1 \qquad h(1) = 1 \qquad |f'(1)| \leq 1 \qquad |h'(1)| \leq 1$$

$$\text{If } f(x^*) = x^* \quad \text{then} \quad |f'(x_m)| > 1 \quad \text{for } m = (1, \ldots, M)$$

$$\text{If } h(w^*) = w^* \quad \text{then} \quad |h'(w_q)| > 1 \quad \text{for } q = (1, \ldots, Q)$$

13 We assume therefore non-negative pay-offs.

14 Throughout, we use emergence to refer to the convergence of the solution within a single generation, whereas evolution refers to the path of the system across generations. If $r \leq 2$ then A^{**} is stable and the system will evolve toward cooperation. If $r > 2$ then A^{**} is unstable and various evolutionary paths are possible.

15 As it turned out, Von Neumann's ideas bear a striking resemblance to what Watson and Crick later discovered about the structure of genetic codes.

16 'Because…the systems encountered in ecology and economics are always open to flows of matter and energy [they] only attain thermodynamic equilibrium at death. Living systems are not equilibrium structures, and all of their important capacities, like adaptation, learning and indeed consciousness spring from this fact…. Living systems are in constant dialogue (not equilibrium) with their environment, and when not visibly evolving maintain the capacity to evolve and change' (Allen 1992: 107).

REFERENCES

Allen, P.M. (1992) 'Modelling evolution and creativity in complex systems', *World Futures,* 34,105 – 23.

Arthur, B.W. (1994) 'Inductive reasoning and bounded rationality', *American Economic Review P&P,* 84, 406 – 11.

Aumann, R. (1981) 'Survey of repeated games', in R.Aumann, Harsanyi, J.C., Hildebrand, W, Maschler, M., Perles, M.A., Rosenmuller, I, Selten, R., Shubik, M. and Thompson, G.L. (eds) *Essays in Game Theory and Mathematical Economics in Honour of Oskar Morgenstern,* Mannheim: Bibliographisches Institut.

Aumann, R. and Sorin, S. (1989) 'Cooperation and bounded recall', *Games and Economic Behaviour,* 1, 5 – 39.

Axelrod, R. (1984) *The Evolution of Cooperation,* New York: Basic Books.

Bicchieri, M.C. (1987) 'Self refuting theories of strategic interaction: a paradox of common knowledge' , working paper, University of Chicago (CERAS).

Binmore, K. (1987) 'Modeling rational players II', *Economics and Philosophy,* 4, 179 – 214.

——(1991) *Fun and Games,* Lexington MA: D.C.Heath.

Binmore, K. and Samuelson, L. (1992) 'Evolutionary stability in repeated games played by finite automata', *Journal of Economic Theory,* 57, 278 – 305.

Dawkins, R. (1976) *The Selfish Gene,* Oxford: Oxford University Press.

Eigen, M. and Winkler, R. (1983) *Laws of the Game: How the Principles of Nature Govern Chance,* New York: Penguin.

Fudenberg, D. and Harris, C. (1992) 'Evolutionary dynamics with aggregate shocks', *Journal of Economic Theory,* 57, 420 – 41.

Fudenberg, D. and Maskin, E. (1990) 'Evolution and cooperation in noisy repeated games', *American Economic Review P&P,* 80, 274 – 9.

Grandmont, J.M. and Malgrange, P. (1986) 'Non linear economic dynamics: introduction', *Journal of Economic Theory,* 40, 3 – 12.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems,* Ann Arbor MI: University of Michigan Press.

Luhmann, N. (1982) *The Differentiation of Society,* New York: Columbia University Press.

May, R.M. (1976) 'Simple mathematical models with very complicated dynamics', *Nature,* 261, 459 – 67.

Maynard Smith, J. (1982) *Evolution and The Theory of Games,* Cambridge: Cambridge University Press.

Papadimitrou, C. (1992) 'On players with a bounded number of states', *Games and Economic Behaviour,* 4, 122 – 31.

Penrose, R. (1989) *The Emperor's New Mind,* New York: Penguin.

Poincarè, H. (1909) *Atti del IV Congresso Internazionale di Matematica,* Roma, 1908, vol. I, Rome: Accademia dei Lincei.

Rapoport, A. (1966) *Two-person Game Theory: The Essential Ideas,* Ann Arbor MI: University of Michigan Press.

Rapoport, A. and Chammah, A.C. (1965) *Prisoner's Dilemma,* Ann Arbor MI: University of Michigan Press.

Rubinstein, A. (1986) 'Finite automata play the repeated Prisoner's Dilemma', *Journal of Economic Theory,* 39, 83 – 96.

Samuelson, P. (1993) 'Altruism as a problem involving group versus individual selection in economics and biology', *American Economic Review P&P,* 83, 143 – 7.

Selten, R. (1975) 'Reexamination of the perfectness concept for equilibrium points in extensive-games', *International Journal of Game Theory,* 4, 25 – 55.

Sen, A.K. (1967) 'Isolation, assurance and the social rate of discount', *Quarterly Journal of Economics,* 81,112 – 25.

——(1985) 'Goals commitment and identity', *Journal of Law, Economics and Organisation,* 1,341 – 55.

Smale, S. (1976) 'Dynamics in general equilibrium theory', *American Economic Review P&P,* 66, 288 – 94.

——(1980) 'The Prisoner's Dilemma and dynamical systems associated to non-cooperative games', *Econometrica,* 48, 1617 – 34.

Von Neumann, J. (1987) *Papers of John Von Neumann on Computing and Computer Theory,* eds W.Aspray and A.Burks, Cambridge MA: MIT Press.

Von Neumann, J. and Morgenstern, O. (1953) *Theory of Games and Economic Behaviour,* 3rd edn, Princeton NJ: Princeton University Press.

Young, H.P. and Foster, D. (1991) 'Cooperation in the short and in the long run', *Games and Economic Behaviour,* 3, 145 – 56.

26
A MODIFIED TREND THROUGH PEAKS APPROACH TO MEASURING POTENTIAL OUTPUT
An application to the Italian economy

Giuliana Campanelli

INTRODUCTION

Potential output is an extremely useful but at the same time quite an ambiguous concept. Over the last thirty-five years many different definitions of potential output have been put forward, and many alternative approaches have been elaborated to quantify those concepts. Every procedure, however, can be open to criticisms for its strong assumptions on how to calculate potential output from figures for actual output, since potential output is not observable. We believe that most of the over-simplified assumptions arise from the impossibility of finding a precise and at the same time manageable measure of potential output for a modern and complex economy (i.e. an economy with many commodities, many types of labour and capital, uncertainty about future demand, variable growth rates over time and in which institutional and sociological factors also play a very important role).

Potential output has very often been defined as 'the maximum' level of output associated with the 'full utilisation' of existing resources. Alternatively, it has meant a 'normal' level of production given 'average' levels of factor utilisation rates, and sometimes has even been interpreted as that level of output which corresponds to some 'optimal' utilisation of productive resources. We should also mention that a recent literature—coming mainly from the IMF (see Adams *et al.* 1987; IMF 1988) and the OECD (see Jarrett and Torrens 1987; Torrens and Martin 1989; and more recently, Giorno *et al.* 1995; Turner 1995)—has defined potential output as 'the maximum level of output consistent with stable inflation'.[1] The inflation constraint is based on the estimate of the natural rate of unemployment at any given point in time. As such, this last concept of potential output isclearly different from the maximum attainable level of output that could be produced with given factors of production, as defined earlier. Indeed, it puts the main emphasis on inflation, with the consequence that the gap between actual and potential output can be viewed more as an indicator of inflationary or deflationary pressure in an economy (depending of course on whether actual

output is above or below potential) rather than as a measure of idle resources existing in that economy.

The scope of this paper is not to review the growing literature on the subject, but to suggest an operative approach to measuring potential output defined along the lines of the maximum feasible capacity output compatible with existing resources. In particular we will propose a straightforward approach which may be regarded as an improvement over the ordinary trend through peaks procedure since it distinguishes between peaks of different capacity.

We will then present the empirical estimates of potential output for the Italian economy (as a whole) over the 1960 – 82 period, based on the approach previously discussed. We took the 1960 – 82 period[2]—i.e. roughly the years from the beginning of the so-called 'Italian miracle' up to those immediately after the second oil shock, since it was crucial for the development of the Italian economy. Over this period, in fact, real GDP grew very rapidly, not only in comparison with past Italian experience but also by international standards: Italy was second only to Japan! As is well known, however, over the sub-period 1973 – 82 growth slowed down and the state of confidence became quite uncertain. The purpose of this second section of the paper is to compare our results with those coming from other sources[3] and to examine whether such a comparison supports our results.

THE APPROACH

Let us first consider the trend through peaks approach as originally elaborated by Klein and Wharton Econometric Forecasting Associates (WEFA) (see Klein and Summers 1966; Klein and Preston 1967). According to this procedure, potential output is assumed to equal production at each cyclical peak, and between peaks it is supposed to grow at a constant rate equal to the slope of the straight line linking the production peaks. A straight line is also extrapolated (using the same slope as the line connecting the previous two peaks) beyond the last one, until later data establish a new peak. This approach—usually used to calculate potential output for manufacturing or for the economy as a whole[4]—is undoubtedly very simple and has the great advantage that it takes into account the characteristics of a real economy (e.g. varying speeds of growth of population, employment and capital stocks) to the extent that these are reflected in GDP peaks. It is open, however, to the following five major criticisms:

1 It always sees potential output at the peaks as an attained level of production, but this means that it really refers to actual output rather than to potential output.[5]
2 It does not distinguish between one peak and another, whereas they can be associated with different levels of unused capacity,[6] and those peaks where

the margin of spare capacity is larger are actually some way below the figure we really want.

3 It is not very reasonable to assume that potential output grows at a constant rate between two peaks, even though they are many years apart. One should expect, indeed, that its growth is somewhat uneven, similar to investment growth which is pro-cyclical (allowing of course for lags in the completion of capital projects).

4 One should not forget that when a new peak has recently occurred the more recent capacity figures should be subject to revisions.

5 It is not able to decompose potential output growth into its main determinants.

Several methods of improving the trend through peaks method have been suggested. In this connection the procedure elaborated by Dhrymes (1976) is worth mentioning. It tries to combine the trend through peaks approach with a production function approach in order to modify the two strong assumptions that growth of potential output is constant between peaks and the current growth of potential output is equal to that obtaining prior to the most recent peak.[7]

In the light of the above mentioned limitations of the historical peak-to-peak approach, we also suggest an alternative procedure, the scope of which is to improve the original peak-to-peak procedure by distinguishing between peaks of different capacity.

The approach we elaborated can be summarised as follows. First we calculate potential output by applying the trend through peaks procedure directly to real GDP. This means that we mark off cyclical peaks in actual output on a log scale and then fit linear segments between successive peaks. This approach, as already emphasised, is relatively simple and, moreover, has the advantage that it takes into account the characteristics of a real economy to the extent that these are reflected in GDP peaks. However, as already discussed, since this trend line might link peaks of actual outputs associated with different degrees of capacity utilisation, we will define it as ‘apparent potential’.

Thus the second stage of our approach consists of calculating—on the basis of the apparent potential—a different potential, which can be called ‘calculated potential’, the peaks of which are taken to reflect the *same minimum level* of unused capacity.

Theoretically, one should consider the unused capacity of both labour and capital, but since figures for unused capital stock for the whole economy are either not available or not very reliable, we choose unemployment as theonly indicator of unused capacity.[8] To measure calculated potential, three steps are required:

1 To select the minimum unemployment rate observed for one peak.[9]

2 To calculate, for each of the other peaks, that level of output which would have been produced if the same labour force percentage associated with the observed minimum level of unemployment was employed. More precisely, the level of output at the 'new' peak for a generic year is calculated, by using Okun's procedure (Okun 1962), as follows:

$$P_t = A_t\,[1 + a/100(U_t - U^*)] \tag{1}$$

where P_t is the level of calculated output associated with the new peak in the generic year *t;* A_t is the level of output associated with the observed peak in the generic year *t;* U_t is the unemployment rate associated with the level of output in the observed peak in the generic year *t;* U^* is the observed minimum unemployment rate in the year when this minimum was recorded; and a is the coefficient linking output and unemployment rate. Later we will apply the long-term coefficient estimated according to Okun's procedure.[10]

3 Once we have calculated all the new peak outputs, we will, once again, fit linear segments between them. Such a trend line through the new peaks is assumed to indicate calculated potential output. Thus calculated potential gives an estimate of that level of output that could have been produced if at each peak demand had increased to employ exactly the same percentage of the labour force associated with the minimum level of unemployment observed in the economy.

It is worth noting that in this context what is really important is not—as the historical peak-to-peak procedure would suggest—whether the peak is the highest point in the neighbourhood, but how much of the peak of actual output, i.e. of apparent potential, is below that of the 'true' potential, i.e. of calculated potential.

As already emphasised, the advantages of this approach relative to the historical trend through peaks procedure is that it is able to distinguish between peaks of different capacity. However, it should still be regarded as an approximation to calculate potential output, at least for the following three reasons:

1 Potential output continues to grow at a constant rate between peaks.
2 One should remember that unemployment is *not* the only indicator of unused capacity over the peaks. And, not having been able to secure figures for the unused stock of capital for the economy as a whole represents quite a serious limitation, since the utilisation of capital may havemoved differently from that of labour, with the consequence that calculated potential might be either 'too low' or 'too high' on this account.
3 Finally—unlike the production function approach—it is not able to decompose movements of potential output into its major determinants. However—as it is widely recognised[11]—the theory underlying the

production function as well as its estimation raises a host of problems. The more important among these are related to the following three aspects:

(a) the specification of the production function;
(b) the meaning as well as the measurement of labour and capital in particular;
(c) the treatment of technical change (often defined as total factor productivity).

THE ESTIMATION RESULTS

We now present the empirical estimates of potential output for the Italian economy based on the approach previously discussed; we took, as already mentioned, the most rapid growth period of its economy: 1960 – 82. The scope of this empirical investigation is to look not only at short-term movements in capacity utilisation, but also at long-run movements in actual and potential output as defined earlier. In the next section we will compare our results with those coming from other sources.

The particular questions we address here are the following:

1 To what extent did the growth of calculated potential output slow down over the second sub-period 1973 – 82?
2 How did the margin of unused capacity—defined as difference between calculated and apparent potential divided by calculated potential—move after 1973?

To answer these questions we first present the results of our estimates of movements in potential output for the Italian economy as a whole, based on our modified trend through peaks approach. Subsequently we report the estimates of potential output coming from other sources; as we already pointed out, the purpose is to see whether such a comparison will increase the confidence in our results.

Estimating Okun's coefficient

We start by estimating Okun's coefficient; subsequently, on the basis of this coefficient, we estimate potential output for the Italian economy as a whole following the modified trend through peaks approach previously suggested.

Three main steps are necessary to obtain Okun's coefficient:

1 First we use a semilog formulation to estimate the unemployment rate *(UR)* as a function of (log of) GDP and a time trend. The relationship was estimated by Ordinary Least Squares (OLS) from annual data for the period

1960 – 82. Different dynamic structures were also tried,[12] and by using the criteria both of goodness of fit and of plausibility of the coefficients, the following equation was chosen as the preferred one.[13]

$$UR = 1.4449 + 0.0089773\,t - 0.00172\,t^2 - 0.139\,LGDP + 0.7195\,UR(-1)$$
$$(0.542) \quad (0.003) \quad (0.0001) \quad (0.053) \quad (0.224)$$

$$R^2 = 0.9051,\ \text{S.E.R.} = 0.00405,\ \text{M.L.L.} = 92.7951$$

$$F_{sc}\,(1,16) = 0.5532,\ F_{ff}(1,16) = 0.3810$$

Where *UR* is unemployment rate (expressed in percentage points); *t* is time trend; *LGDP* is log of GDP (at factor cost expressed at 1970 prices); and *UR* (−1) is lagged variable for unemployment rate. The short-term coefficient linking unemployment rate and output suggests that, at any point in time, for each extra one per cent of GDP the unemployment rate is 0.139 points lower.

2 The second step consists of calculating the long-run coefficient linking unemployment rate and output on the basis of the estimated short-run coefficients. If we write the previous equation in the following general form:

$$UR = a + \beta_1 t + \beta_2 t^2 + \beta_3 LGDP + \beta_4 UR(-1)$$

the long-run coefficient between *UR* and *LGDP* can be obtained as follows:[14]

$$\beta_5 = \frac{\beta_3}{(1 - \beta 4)} = -\frac{0.139}{1 - 0.7195} = -0.4955$$

It clearly emerges that the long-term responsiveness of unemployment rate to output is more than three times that shown for the short period.

3 The third step consists of obtaining the long-run Okun's coefficient implied by the above estimation. This is simply the inverse of the long-run coefficient of GDP, i.e. $1/\beta_5$.

As a result in our case the long-run coefficient *a* linking LGDP and *UR* would be given by the following expression:

$$a = \frac{1}{\beta_5} = -\frac{1}{-0.4955} = -2.018$$

This implies that, in the long run, an addition of one percentage point in the unemployment rate is associated with a loss of nearly 2 per cent in output.

Estimating potential output

The results of the modified trend through peaks methodology to calculate potential output are presented in Table 26.1.[15] A few clarifications are necessary to understand this table i.e.,

(a) The first year peak 1963 is associated with the minimum level of unemployment, i.e. 3.9 per cent. This implies that in this year apparent and calculated potential coincide; and
(b) Calculated potential for the other peaks has been obtained by applying Equation (1) to the corresponding GDP peak figures.

More specifically, substituting the figures for a (i.e. the long-run Okun's coefficient equal to −2.013), and considering that the minimum level of unemployment over the 1960 – 82 period for the Italian economy was 3.9 per cent, Equation (1) can be rewritten as follows:

$$P_t = A_t(1 + 0.0202(U_t - 3.9)) \qquad (2)$$

In addition, Figure 26.1 provides a straightforward graphical representation of our results (based on annual figures expressed in ratio scale).

Some preliminary conclusions

Looking at the results based on our modified trend through peaks approach, we are now in a position to answer the two questions posed earlier on the evolution of potential output for the Italian economy. It is easy to see that the approach implies that:

1 the growth of potential output becomes lower through time, particularly after 1973.

Table 26.1 Actual output, apparent potential and calculated potential based on modified trend through peaks approach, 1960 – 80

	(a) Years of GDP peaks	*(b) Apparent[a] potential*	*(c) Unemployment as % of total labour*	*(d) Calculated[b] potential*
	1963	39,859	3.9	39,859
	1970	56,791	5.4	58,510
	1974	66,310	5.4	68,317
	1980	76,873	7.6	82,613
Comp. growth rate per annum	63–74	4.74	—	5.26
	74–80	2.49	—	3.22

Source: Figures calculated from ISTAT, National Accounting.
Notes:
Figures at 1970 prices, billion Lire
[a]This column shows the 'peaks' of real GDP at 1970 prices
[b]It has been obtained as follows:

$$da_i = ba_i \left[1 + \frac{a}{100}\,(ca_i - c^*_{63})\right]$$

where a_i are years of GDP peaks and a is the estimated long run Okun's coefficient equal to −2.018

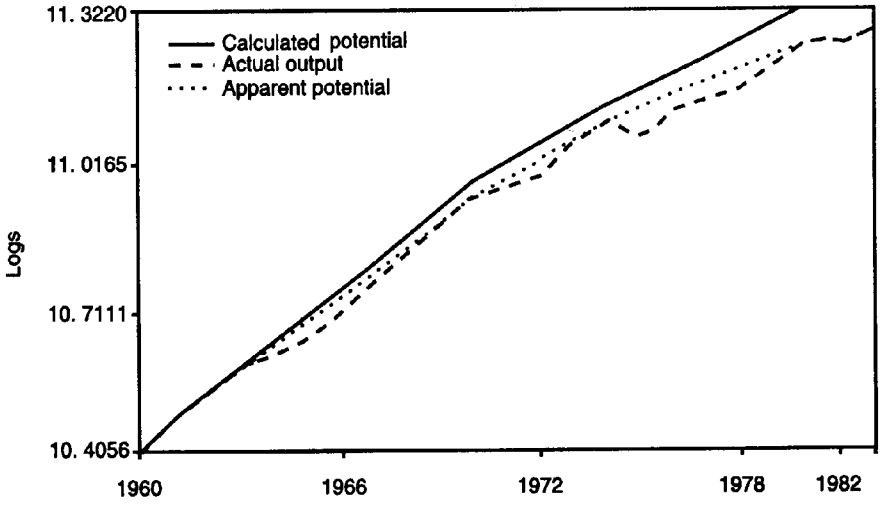

Figure 26.1 Actual output, apparent potential and calculated potential based on modified trend through peaks approach, 1960 – 82

2 The rate of growth of calculated potential, however, slowed down less than the growth of apparent potential: a reflection of higher levels of unemployment in 1980 associated with a lower demand.
3 In spite of the slowdown of potential output growth, the margin of unused capacity—measured as difference between calculated and apparent potential divided by calculated potential—increased substantially over the second sub-period 1974 – 82. This is due to the fact that 1980 GDP peak is associated with a much higher level of unemployment than the other peaks.

A COMPARISON WITH OTHER ESTIMATES

Let us now summarise the major findings on the evolution of potential output and the margin of unused capacity for the Italian economy resulting from other studies.

As already emphasised, we want to see whether a comparison with other estimates—even though based on different approaches—will increase confidence in our results, especially as the much simpler method presented above does not explicitly take account of variables such as the growth of the capital stock.

Before presenting these results we should mention two problems that arise from the comparison: first, the figures do not always refer to the same time periods; and second, some of them only apply to manufacturing or industry rather than to the economy as a whole. We believe, however, that these do not constitute severe limitations for two main reasons:

(a) In spite of the differences in the periods analysed, all the estimates give us information about movements in potential output after 1973 *vis-à-vis* the 1960s.
(b) Industry, and manufacturing in particular, is an extremely important sector of the Italian economy and, as national accounting figures can confirm, it was the major contributor to the growth of total value added over the whole period considered.

Let us briefly recall the major sources used and their different approaches. First, the Bank of Italy provides figures for potential output, based on the original trend through peaks approach; in particular we refer to the estimate of potential output elaborated by F.L.Signorini (1986) for the period 1954 – 85. The method used by the author is the same as that previously applied by G.Bodo (1981), which consists of calculating the indices of potential production first at industry group level, and then aggregating[16] them to obtain indices for total industry and services. The main difference, however, concerns the level of disaggregation used as a base forthe general index; Signorini's method calculates this index from more than fifty manufacturing industry groups whereas the previous method used thirty-five only. Thus it is supposed to increase the reliability of the results.

Second, the Italian National Institute for the Study of Economic Trends (ISCO) —as already mentioned[17]—produces figures for used capacity in plants in industry (excluding mining). The figures for the 1969 – 80 period are based on the survey of the firms' method.

Third, the estimates of potential output for the Italian economy based on a production function approach have been elaborated mainly by the research departments of the IMF and the OECD. Here we only consider the results of the work of Turner (1987),[18] who estimates potential output for Italian manufacturing over the period 1965 – 86, following the 'engineering'[19] approach previously used by the IMF.[20] The estimated effect of the oil price shocks on

capital stock growth rates and on potential output growth rates is an essential element of this approach. Moreover, the work of Turner may be regarded as interesting for an additional reason: it tries to break down potential output growth into its major determinants, i.e. technical change (or total factor productivity), labour force, capital stock and mean age of capital stocks. We do not include the results of these estimations, however, for the simple reason that here we are only interested in movements of actual and potential output as a whole.

By looking at the Bank of Italy, ISCO and Turner's estimates it seems that, in spite of the divergent approach used, they all arrive at conclusions which are quite similar to those found by applying our modified trend through peaks approach, i.e.

1 The growth of potential output slowed down significantly after 1973.
2 The margin of unused capacity increased after 1973, in spite of the slowdown in potential output growth. In particular, the 1980 GDP peak is associated with a higher level of unemployed resources *vis-à-vis* the other peaks. The only difference is with the Bank of Italy estimates since, being based on the original trend through peaks approach, they associate at each peak at industry group level the same margin of unused capacity.

However, if we take into account the relatively recent IMF and OECD literature on potential output based on the natural rate of unemployment, we will *not* necessarily arrive at similar conclusions. In particular, the first result concerning the slower growth of potential output after 1973 continues to hold true, even though the two sources give very different figures for its extent: the reduction in the growth rate of potential output between the two sub-periods 1966 – 73 and 1974 – 9 is −3.7 percentage points according to the IMF estimates (from 5.1 to 1. 4 per cent), against −1.2 percentage points according to the OECD estimates (from 3.8 to 2.6 per cent).

The second result, however, relating to the increasing margin of unused capacity, no longer appears to be valid. On the contrary, according to the non-accelerating inflation potential output approach, after the beginning of the 1970s actual output in Italy (as well as in most of the major industrial countries) was very often above potential. As a result the output gap—measured as actual minus potential divided by potential—was positive and became increasingly so until the late 1970s. It is only in the early 1980s that—according to this literature—actual output fell below potential.

The divergence in the results between the IMF and OECD sources mainly reflects the enormous discretion involved in the estimates of the natural rate of unemployment defined as 'that rate of unemployment which would prevail in equilibrium when expectations about inflation are realised' (see Adams *et al.* 1987). The explanation for the difference between the IMF/OECD and our results lies in the different definitions of potential output and, in particular, in the

different assumptions about the 'full employment rate of unemployment'. The conclusion that the gap became increasingly positive over the 1970s, in spite of the high rate of unemployment actually recorded, stems from the finding of the more recent IMF and OECD estimates that the natural rates of unemployment have generally increased significantly over the second sub-period.[21] This was largely attributable

> to the interaction of labour market rigidities with a series of adverse disturbances—such as the oil price increases—that called for a moderation of real wage gains if high employment was to be maintained. However because of real wage rigidity in Italy as well as in many other industrialised European countries, equilibrium levels of unemployment were lower, natural rates of unemployment rose and potential output in these countries declined below the levels that could have been achieved with more flexible labour market.
>
> (Adams *et al.* 1987:26)

As a result, the non-accelerating inflation potential output turned out to be much lower than the one calculated in 'engineering' terms, particularly if it is based, as in the case of our estimates, on the minimum level of unemployment actually recorded.

CONCLUDING OBSERVATIONS

In this paper a straightforward approach to measuring potential output and the margin of unused capacity was developed, which should be regarded as an improvement on the ordinary trend through peaks procedure because it distinguishes between peaks of different capacity. The application of thisapproach necessitated the previous estimation of the long-run coefficient linking output and unemployment rate which was obtained by applying a direct procedure, as indicated by Okun. We then presented the result of the empirical estimates of potential output for the Italian economy based on the approach discussed, and compare them with those coming from other sources.

The main findings on the evolution of potential output for the Italian economy over the 1960 – 79 period are the following:

(a) the growth of potential output slowed down substantially after 1973: a conclusion confirmed by all the estimates in spite of the substantial divergences in the concepts and in the approaches used. Our procedure also shows that the rate of growth of calculated potential slowed down (after 1973) less than the growth of apparent potential: a reflection of higher levels of unemployment in 1980 associated with a lower demand.

(b) Conclusions regarding movements in the margin of unused capacity vary, depending on the definition of potential output used and in particular on the different assumptions regarding the full employment rate of unemployment. The estimates of potential output based on the minimum level of unemployment actually recorded, like the one we proposed, indicate an increasing margin of unused capacity after 1973 *even at peaks*—a result supported by other estimates that are only concerned with the measurement of idle resources existing in the economy. The estimates based on the natural rate of unemployment find, on the contrary, that actual output in Italy was very often above potential over the same years, even though actual unemployment was quite high.

ACKNOWLEDGEMENTS

The basic ideas of this paper were elaborated while I was a Ph.D. student at Cambridge University. Among the people who assisted in shaping them it is to my supervisor W.Brian Reddaway that I owe the greatest debt. Ken Coutts and Hossein Samiei provided very helpful comments; I would like to thank them too.

NOTES

1 The production function non-accelerating inflation approach to potential output was originally formulated by Perloff and Wachter. See Perloff and Wachter 1979. The most recent contributions along the same line can be found in Adams and Coe (1990) and Kuttner (1992,1994).

2 We took the 1960 as base year to measure growth mainly for statistical reasons: a series of (supposedly) consistent national accounting figures goes back to 1960. In the event, however, it appears to be a convenient post-Treaty-of-Rome starting point. We stopped at 1982 rather than 1979 (the year of the second oil shock) since later data for GDP show that 1979 was not a peak year. We did not include the other 1980s figures because this period requires a separate investigation, since the growth of Italian GDP became much more unstable to the point that Italy has never fully recovered.

3 We also include those results for the Italian economy deriving from the literature on potential output based on the natural rate of unemployment.

4 The indices of potential production can be either calculated directly from aggregate figures, or first computed at industry level and then aggregated to obtain indices for total industry and services. However, many problems arise from the aggregation procedure, usually based on value added weights, since industries do not generally reach a major peak at the same time.

5 To quote the 'inventors' of this method, potential output of an industry at a particular time is 'the maximum sustainable level of output that the industry can attain within a very short time *if the demand* for its products *was not a constraining factor,* when the industry is operating with its existing stock of capital at its customary level of intensity' (Klein and Summers 1966:2, emphasis added).

6 A few years after the appearance of the peak-to-peak approach 'the inventors' themselves admitted this weak point: 'peaks may be marked off as capacity utilization peaks when in fact there may have been considerable underutilisation of capacity' (Klein and Preston 1967:34).
7 The advantage of this modified trend through peaks procedure is 'that it makes more efficient use of available information regarding what happens to potential output between peaks and, most important, what has been happening since the most recent peak'(Christiano 1981:156).
8 For example, in the case of Italy the National Institute for the Study of Economic Trends (ISCO) only produces figures for used capacity in plants in industry (excluding mining). In addition the figures are based on the survey of the firms' method, i.e. on direct interviews to entrepreneurs about the degree of utilisation in plants in the last three months in relation to the maximum utilisation rate. This means that one should be very cautious in interpreting the figures for many reasons. First, they are based on the subjective judgements of different industrialists with the result that they are not truly comparable. Second, they look at capacity utilisation from the demand side only, whereas one should look at movements in capacity utilisation from the supply side as well. Third, they refer to a short-term period. Finally, the degree of used capacity in plants could in itself be deceptive in judging idle capacity in an economy, since it is perfectly conceivable that in a period of slow growth of output, many firms could decide to increase the degree of utilisation of the existing plants rather than to reduce it.
9 This figure would include both voluntary as well as frictional unemployment, which are not easy to separate in a real economy. All the unemployment rates above the minimum are assumed to reflect involuntary unemployment.
10 See next section. One can also try to estimate the long-term coefficient by using a more indirect method as suggested by Godley and Shepherd (1964).
11 On this issue see Harcourt 1972, particularly Chapter 2.
12 We tried different alternative specifications in order to detect whether or not relevant variables were omitted from the model. The addition of a lagged values of GDP (one and two years) and further lags of unemployment rates (two and threeyears) did not prove to be significant. The relevant parameters, however, remained quite stable.
13 The figures in brackets are standard errors, R^2 is the multiple correlation coefficient adjusted for the degree of freedom, S.E.R. is the standard error of the regression, M.L.L. is the maximised values of the log-likelihood function. The diagnostic tests reported, all provided by DATA-FIT (1987), are F_{sc}—a Lagrange multiplier test of serial correlation—and *Fff*—the Ramsey reset test for functional form.
14 Here we are assuming a steady-state growth in unemployment rate, i.e. $UR = UR(-1) = UR^*$.
15 We took GDP expressed at 1970 prices and used annual figures.
16 The weights used for the aggregation are based on figures for industrial production. It is worth adding that the potential output elaborated by the Bank of Italy does not correspond to our apparent potential, since each peak is not associated with zero margin of unused capacity. This outcome is due to aggregation problems since the different industry groups considered did not reach a major peak at the same time.
17 See note 8 above.
18 We discuss only Turner's 1987 work for two main reasons:

(a) It represents an updated version of the two papers previously written by Artus (1977) and Artus and Turner (1978).

(b) Two interesting refinements have been introduced relative to the methodology previously employed. First, figures for the stock of capital have been adjusted for obsolescence that occurred in response to the marked increases in the relative price of energy in 1973 – 4 and 1979 – 80 respectively; second, in the empirical analysis, allowances were made for shifts in the trend element; in particular, three separate trends were employed.

19 The term 'engineering' is used to distinguish this approach based on the 'structural unemployment rate' from that based on the natural rate of unemployment.

20 See in particular Artus 1977; Artus and Turner 1978. Later we will also discuss the results in the light of the subsequent estimates of the IMF and the OECD based on a different concept of potential output, i.e. the maximum level of output consistent with stable inflation.

21 'Actual and natural rates of unemployment have been similar, notwithstanding extraordinary high unemployment rates' (Adams *et al.* 1987:15).

REFERENCES

Adams, C. and Coe, D.T. (1990) 'A system approach to estimate the natural rate of unemployment and potential output for the United States', *IMF Staff Papers,* 37, 2: 232 – 93.

Adams, C., Fenton, P.R. and Larsen, F. (1987) 'Potential output in major industrial countries', *Staff Studies for the World Economic Outlook,* IMF, August: 1 – 38.

Artus, J.R. (1977) 'Measures of potential output in manufacturing for eight industrial countries, 1955 – 78', *IMF Staff Papers,* 24, 1:1 – 35.

Artus, J.R. and Turner, A.G. (1978) 'Measures of potential output in manufacturing for ten industrial countries, 1955 – 80, IMF, unpublished.

Bodo, G. (1981) 'Misura ed analisi della capacità produttiva', *Banca d'ltalia, Bollettino,* special issue.

Christiano, L.J. (1981) 'A survey of measures of capacity utilisation', *IMF Staff Papers,* 28,1:144 – 98.

Data-Fit (1987) *Interactive Econometric Program,* by M.H.Pesaran and B.Pesaran, Oxford: Oxford University Press.

Dhrymes, P.J. (1976) 'Impacts of an overvalued currency on domestic income, employment and prices', Department of Economics, Columbia University, New York, unpublished.

Giorno, C, Richardson, P., Roseveare, D. and Van Den Noord, P. (1995) *Estimating Potential Output, Output Gaps and Structural Budget Balances,* Working Paper no. 152, Paris: OECD.

Godley, W.A.H. and Shepherd, J.R. (1964) 'Long term growth and short term policy', *National Institute Economic Review,* 29:26 – 38.

Gordon, J.R. (1979) 'A comment on the Perloff and Wachter paper', in K.Brunner and A.H.Meltzer (eds) *Three Aspects of Policy and Policy Making: Knowledge, Data and Institutions,* Amsterdam: North Holland.

Harcourt, G.C. (1972) *Some Cambridge Controversies in the Theory of Capital,* Cambridge: Cambridge University Press.

IMF (1988) 'Potential output and capacity utilisation in major industrial countries', *World Economic Outlook,* October, Annex II: 45 – 8.

Jarret, P. and Torrens, R. (1987) *A Revised Supply Block for the Major Seven Countries in Interlink,* Working Paper no. 41, Paris: OECD.

Klein, L.R. and Preston, R.S. (1967) 'Some new results in the measurement of capital utilisation', *American Economic Review,* 57, 1:34 – 57.

Klein, L.R. and Summers, R. (1966) *The Wharton Index of Capacity Utilisation,* Studies in Quantitative Economics no. 1, Philadelphia PA: University of Pennsylvania.

Kuttner, K.N. (1992) 'Monetary policy with uncertain estimates of potential output', *Economic Perspectives,* 16, 1:2 – 15.

——(1994) 'Estimating potential output as a latent variable', *Journal of Business and Economic Statistics,* 12, 3:361 – 8.

Okun, A.M. (1962) 'Potential GNP: its measurement and significance', American Statistical Association, *Proceedings of the Business and Economics Section,* 1, 98 – 104.

Perloff, J.M. and Wachter, M.L. (1979) 'A production-function-non-accelerating-inflation approach to potential output: is measured potential output too high?' in K.Brunner and A.H.Meltzer (eds) *Three Aspects of Policy and Policy Making: Knowledge, Data and Institutions,* Amsterdam: North Holland.

Signorini, L.F. (1986) *Nuove valutazioni della capacità utilizzata in Italia,* Servizio Studi Banca d'ltalia, Temi di discussione, no 81, Rome.

Torrens, R. and Martin, J.P. (1989) *Potential Output in the Seven Major OECD Countries,* Working Paper no. 66, Paris: OECD.

Turner, A.G. (1987) 'Potential output in manufacturing for ten industrial countries: revised estimates, 1965 – 86', IMF, unpublished.

Turner, D. (1995) *Speed Limit and Asymmetric Inflation Effects from the Output Gap in the Major Seven Economies,* Economic Studies no. 24, Paris: OECD.

27
THE WAGE-WAGE SPIRAL AND WAGE DETERMINATION
An empirical appraisal for Italy

Sergio Destefanis

INTRODUCTION

Throughout the years, a feature that has characterised Geoff Harcourt's economic policy analysis has been his careful consideration of the structural features of the inflationary process in postwar capitalist economies.[1] In fact, the central insight of Harcourt's analysis of inflation is that the latter cannot be seen simply as the result of a gap between aggregate demand and supply (the 'demand-pull' view), or of a sequence of exogenous supply shocks (the 'cost-push' view), or even as a combination of these two factors. Indeed, while both sets of elements may be very important in given circumstances, a complete explanation of the postwar inflationary process in advanced capitalist economies must rely on the consideration of the institutional aspects of these economies. More precisely, the mainspring of that process is to be found in the conflict around the distribution of income across various social groups. In this view, inflation is the result of a clash between non-compatible claims to the social product, and its pace ultimately depends on the degree of consensus across socioeconomic groups, the weight they attach to considerations of fairness and the bargaining power they wield. Since these elements are largely determined by social and institutional factors, no analysis of the inflationary process can be complete if it does not allow for the latter factors.

In considering how to validate this view empirically, no exhaustive testing can be carried out without taking into account social and institutional elements, something which might prove extremely difficult due to evident difficulties of definition and measurement. This paper proposes an indirect and simpler approach, putting to test the implications of the above view. If wage settlements are chiefly determined by the desire to achieve a fair distribution of the social product, their main aims should be to compensate forand even anticipate rises in the cost of living, and 'of both protecting established (and hallowed) relativities and enabling wage-earners to share in the trend rise in overall national productivity' (Harcourt 1974:522). As will be shown below, a crucial implication

of these efforts is that cross-industry wage comparisons matter in sectoral wage determination.

Accordingly, the weight of a 'fair wage' norm and of bargaining power in wage determination will be assessed by estimating the empirical relevance of cross-industry wage comparisons *vis-à-vis* more traditional determinants of wages (for instance, sectoral labour productivity and the rate of unemployment). For this purpose, after considering in some detail the concept of cross-industry wage comparisons, a sectoral wage equation is specified, allowing for various measures of outside wages among the regressors. This equation is then estimated for twenty-one industries of the Italian economy, using annual data spanning the 1951 – 93 period. The Italian economy has been chosen as a field of inquiry in order to test the relevance of cross-industry wage interactions in a context where evidence has been far from conclusive. Evidence which favours the relevance of cross-industry wage comparisons for sectoral wage determination in the Italian economy would be novel and quite unexpected *a priori*.

The next section examines various specifications of the link between wage relativities and sectoral wage determination. Particular attention is paid to the concepts of wage leadership and wage contour, and their relevance for the definition of outside wages. The third section sets out an empirical framework that allows for the existence of these interactions in sectoral wage determination. Within this framework it will be possible to assess the importance of cross-industry interactions with respect to other determinants of wages, as well as the relative explanatory power of different measures of outside wages. After a brief introduction to the dataset, the following section, devoted to the Italian case, presents the results of sectoral wage estimates. Various measures of outside wage are considered and their performance is compared. The concluding section comments on these results for the process of wage determination in the Italian economy and on more general considerations that can be drawn from the present exercise.

CROSS-INDUSTRY WAGE COMPARISONS AND THE INFLATIONARY PROCESS

Wage relativities and wage determination

The relevance of wage relativities for the determination of money wages has always been of paramount importance in post-Keynesian economics. Indeed, in his recent and thoughtful survey, Lavoie (1992: ch. 7) emphasisesthat conflict over the distribution of income, originated by requirements of equity and fairness, is the core of the post-Keynesian theory of wage inflation. More particularly, wage claims are influenced by workers' perceptions of a fair relative wage, of a fair real wage and of a fair income share. In other words, workers will press for rises in wages whenever they feel that their rightful position relative to other worker

groups is disrupted, or that their purchasing power is unduly eroded, or that their share of income is too low with respect to the profit share. While it is clear that wage relativities are directly related to the first factor (fair relative wages), it is perhaps less easy to realise that workers' perception of other groups' wages also matters insofar as fairness in real wages and income shares are concerned.

As a matter of fact, fairness is very largely a relative concept, and wage settlements in other industries are likely to be a very important indicator of what is a fair wage claim in a particular sector. In other words,

> comparisons are required. These comparisons, in turn, require information about the profits of the firm…the profits of the industry, the structure of wages, the deals negotiated elsewhere, the value of marginal benefits obtained elsewhere, the recent trend in earnings and so on.
>
> (Lavoie 1992:381)

Yet, to acquire and process information about all these variables is costly, meaning that

> even if information could be accurate and comprehensive, only part of it can be processed by the union leaders and the members of the opposing negotiating team…. This explains why the wage increases arrived at in some key industries often have such an impact on all sectors of the economy.
>
> (Lavoie 1992:382)

Now, this quotation makes it clear not only that outside wages are the element of comparison most likely to be held in mind by workers in the course of wage negotiations, but also that wages in some particular sectors may be prominent in the perception of other workers' groups. Indeed, post-Keynesian analysis takes from institutionalist labour economics the idea of the existence of some key sectors, whose wages dictate the evolution of wages in the other sectors. However, opinions on the nature of these key sectors vary with different authors, making it appropriate to consider this matter in greater detail below.

Finally, note that in recent years the concept of cross-industry wage comparisons has been more and more utilised also in the mainstream analysis of the labour market. It will be enough to mention in this respect threestrands of mainstream labour economics where the specification of alternative earning opportunities is of paramount importance for wage determination: job-search theory, efficiency-wage models and union or bargaining models. In all these approaches, alternative earning opportunities usually depend upon alternative (or outside) wages, as well as on the probability of finding another job and on unemployment benefits. Accordingly, the links between the empirical implications of these models and those of post-Keynesian theory will be further analysed in the next section.

The concept of outside wages in cross-industry wage comparisons

Throughout the years, institutionalist labour economists and then post-Keynesians have devoted considerable attention to the mechanism through which rises in wages are transmitted from one industry to the other. Authors who have reviewed this body of literature[2] agree that two main views are discernible. According to the first one, the rate of wage inflation is determined by inconsistencies among the pay norms of different groups, or between these norms and the other economic factors determining wage relativities (for instance scarcities of particular types of workforce). In the purest version of this approach, the *generalised spillover hypothesis* mainly developed by Wood (1978), wage settlements in an industry depend on past wage settlements in all other industries. This rather extreme prediction can be, as it were, scaled down by considering what Dunlop (1944) defines as a *wage contour*.

According to Dunlop, workers in a given industry are not likely to care or be informed about wages of all other industries, but rather about wages of contiguous industries. Contiguity may be defined spatially, or in economic terms, such as the intensity of input-output relationships. From the empirical standpoint, whereas the generalised spillover hypothesis postulates that, for any given sector, outside wages are equated to average wages for the rest of the economy, this scaled-down version (which can be called for convenience 'bounded spillover hypothesis') takes instead average wages for the rest of the wage contour.

Let us now turn to the second main approach to the wage-wage spiral. A group (or a subset of groups) of workers sets the rate of growth in money wages and is followed by the other groups which try to maintain the previous relativities. This is the *wage leadership hypothesis* pioneered by Dunlop (1944) and Ross (1948), where wage settlements in any industry depend on past wage settlements in the leader (or key) industry. When the wage leadership hypothesis is considered empirically, the main issue is which industry (or industries) should be considered as leader. There are various views in the literature on this matter, which can be described as follows.

According to Kaldor (1959:294), wage leadership is taken by the industrywhich shows the fastest growth in labour productivity over a given period. Indeed, in Kaldor's view, this industry should also be the one showing the largest rises in wages, as a large proportion of the growth in productivity goes into higher wages, and not into higher profits or lower prices. Another classic view is expounded in Eichner (1976:159). Here the wage leader is found in the industry of highest economic importance for the country, or *bellwether industry*. The rationale is that once trade unions have secured a deal in this sector, from what are arguably the most powerful corporations in the country, they are not likely to ask for less favourable terms in other sectors of the economy. Also, the bellwether industry usually enjoys the greatest attention from

the media, favouring the diffusion of information about its wage settlements. Finally, another view which has enjoyed some attention in the literature[3] is that wage leadership has been taken, at least from the 1960s, by the public sector. In the last decades, this sector has become heavily unionised throughout the western world and, unlike the private sector, does not possess obvious and immediate checks against very high wage claims. This, plus the fact that some public services such as education and health undergo a very close scrutiny from public opinion, has created in this sector *a priori* favourable conditions for wage leadership.

To sum up, wage leadership is likely to fall upon a sector which over a given period consistently awards large rises in wages. On economic grounds, workers in other sectors should consider a number of variables related to this sector (wages, but also, for instance, sales, profits and productivity). Yet considerations of fairness and of diffusion of information make it likely that wages alone are going to be considered among these variables, and that the rate of growth in wages negotiated by the key industry is going to set the standard for a fair and acceptable rate of growth in wages across all industries. However, note that the notion of wage contour has been elaborated within this very approach. Hence it is possible that more than one wage leader can exist in an economy, each of them dictating the pace of wage inflation within its own contour.

Authors who have considered the empirical relevance of cross-industry wage comparisons[4] have usually taken the wage leadership hypothesis to be more reasonable and deserving of attention. Accordingly, this is the hypothesis that will be examined in the empirical analysis. Furthermore, some attention will also be devoted to the possible existence of wage contours, following the simple approach proposed by Sargan (1971).

WAGE RELATIVITIES AND WAGE DETERMINATION: AN EMPIRICAL FRAMEWORK

A baseline model

This section proposes an empirical framework for the analysis of cross-industry interactions in sectoral wage determination. The main aim of this exercise is to shed light on the importance of cross-industry interactions with respect to other determinants of wages throughout the economy, and on the relative explanatory power of different measures of outside wages. Hence the wage equation utilised for purposes of estimation and testing will be specified in a manner consistent with different approaches to wage determination and cross-industry wage comparisons. Consider the following expression:

$$W_{it} = w\,[W_{at}/C_t,\ F_{it}\ ,\ P_{it}/C_t,\ B_t/C_t,\ U_t\ ,\ Z_t\ ]\ C^e_t \qquad (1)$$

where W_{it} are money wages in industry i, W_{at} is a measure of outside (or alternative) wages, C_i are consumer prices, F_{it} is a portmanteau term for the variables determining the real product wages claimed by firms[5] (such as labour productivity), P_{it} are producer prices in industry i, B_t are unemployment benefits, U_t is the rate of unemployment, and Z_t is a portmanteau term for any residual determinants of bargaining power and of the wage norm, such as taxation or the state of labour legislation. Superscript ^{e} denotes expected values.

Equation (1) basically stems from a bargaining approach and allows for all the determinants of wages that have been discussed in the previous section. The presence of $C^e{}_t$ means that workers strive for a (consumer) real wage target, but can only do so conditionally on their *expectations* for consumer prices. The term F_{it} accounts for the evolution of the profit share within the industry, while role and specification of W_{at}/C_t have been treated in the previous section. Both variables are expected to enter (1) with a positive sign. Variables P_{it}/C_t, B_t/C_t, U_t and Z_t are essentially related to considerations of bargaining power. The coefficient attached to the gap between producer and consumer prices is expected to be bounded between zero and one, and to be larger the weaker is workers' bargaining power. The latter is also expected to be directly related to unemployment benefits (in terms of consumer goods) and inversely related to the rate of unemployment.

It is easy to show[6] that specifications very similar to (1) can be obtained within formal bargaining models, efficiency-wage models or even within competitive labour market models. As already recalled, in all these approaches alternative earning opportunities depend on alternative (or outside) wages, alongside with the probability of finding another job and unemployment benefits. However, in these models alternative wages are usually equated to a notion of 'aggregate' or 'average' wage, whose empirical treatment is far from straightforward.[7]

For instance, if one assumes the existence of model-consistent expectations, in order to determine wages in any given industry, workers must form expectations about aggregate wages, which in turn are a weighted average of sectoral wages. A possible solution to this problem of infinite regress is to posit the existence of 'consensus expectations' about aggregate wages. Now, in the presence of significant costs of collection and processing of information, workers might well resort to wages in one or few key sectors as the basis for these 'consensus expectations'. As a consequence, it is worth investigating the wage leadership hypothesis as a mechanism of cross-industry wage comparisons, even if the theoretical grounds for (1) differ from the post-Keynesian views expounded in the previous section.

In this sense, a problem of observational equivalence might be said to exist in the present context. Yet even in this case one can gather indirect evidence about the weight of a 'fair wage' norm and of bargaining power in wage determination by assessing how the influence of wage relativities varies across different industries. If wages of the key sector were just a proxy for expectations of

'aggregate wages', their influence on sectoral wage determination should be rather pervasive. On the other hand, if their significance stems from the power to enforce a 'fair wage' norm, the strength of wage-wage interactions should depend on the latter factor, and these interactions are likely to be stronger in industries belonging to the so-called core (or primary) segment of the labour market.

The estimated equations

In order to see how (1) can be utilised in empirical work, consider first the following expression:

$$W_{it} = w[W_{at}/C_t, X'_{it}, U_t]\, C^e_t \qquad (2)$$

The above formula is a simplification of Equation (1) essentially deriving from considerations of empirical expediency. As a matter of fact, the size of the available sample (forty-three annual observations) makes parsimony of specification a rather high priority. Proper sectoral measures of taxation are not readily available, suggesting that Z_t be excluded from the equation. Indeed, the ratio between (economy-wide) direct taxation and nominal GDP was tried in preliminary estimates without any success, and was retained only as part of the set of instrumental variables. Term F_{it} is represented exclusively through X'_{it}, that is the output per person in industry i, X_{it}, multiplied by the relative price P_{it}/C_t. The implicit assumption of equal coefficients for these two variables was tested with success in Pesaran and Lee (1993), and its adoption here in preliminary estimates yielded a much higher fit than the other simple assumption of a zero coefficient for the relative price. For short, variable X'_{it} will be referred to as 'productivity'. Finally, the exclusion of B_t/C_t rests on more solid grounds, as a proper system of unemployment benefits virtually does not exist in Italy.

In order to derive an estimable equation from (2), consider a steady state where $C^e_t = C_t$, and assume that in this situation the functional link among the variables singled out in (2) can be represented through the following log-linear equation:

$$w_{it} - c_t = a_0 + a_1\omega_{at} + a_2 x'_{it} - a_3\, u_t \qquad (3)$$

where lowercase letters denote natural logarithms of the respective variables, ω_{at} stands for $w_{at} - c_t$, and the signs of the coefficients follow from the discussion about (1). Departures from this steady state can be modelled through an adaptation of the error correction mechanism pioneered by Sargan (1964). In the post-Keynesian literature, cross-industry wage interactions are usually couched in terms of rates of change alone. However, there is no *a priori* reason why workers should not attempt recovering the steady-state *levels* of relativities, as well as of purchasing power and of income shares, once these have disrupted. Neglecting to model this mechanism would simply amount to the loss of

potentially useful information in the data. As a consequence, log differences in wages are made conditional on the expected log differences of consumer prices, productivity and outside wages, and on the lagged level gap between consumer real wages, productivity and outside wages, as well as on the level of (the log of) the rate of unemployment.[8] Using the difference operator Δ, one can write:

$$\Delta w_{it} = b_0 + b_1 \Delta c^e_t + b_2 \Delta \omega^e_{at} + b_3 \Delta x'^e_{it} - b_4(w_{it-1} - c_{t-1}) + b_5 \omega_{at-1} + b_6 x'_{it-1} - b_7 u_t + e_t \quad (4)$$

where e_t is a white-noise error term. The simplest assumption that can be made about expectations is that they are model-consistent, implying that actual and expected values only differ through a white-noise term orthogonal to the variables included in (4). Then, actual log differences are substituted to the expected ones, and the equation can be written as:

$$\Delta w_{it} = b_0 + b_1 \Delta c_t + b_2 \Delta \omega_{at} + b_3 \Delta x'_{it} - b_4(w_{it-1} - c_{t-1}) + b_5 \omega_{at-1} + b_6 x'_{it-1} + - b_7 u_t + e'_t \quad (4')$$

where e'_t is the new error term. Equation (4') can be estimated consistently through Instrumental Variables. Given that some of the variables chosen as regressors ($\Delta x'_{it}$ in particular) are likely to be influenced by Δw_{it}, the adop tion of Instrumental Variables would be in any case preferable. However, the use of this procedure is bound to decrease the efficiency of the estimates, making it advisable to reparameterise the equation in terms of variables which are likely to be less collinear. A natural solution in this sense is to re-specify real wages in terms of deviations from productivity. This yields:

$$\Delta \omega'_{it} = \beta_0 + \beta_1 \Delta c_t + \beta_2 \Delta \omega'_{at} + \beta_3 \Delta x'_{it} - \beta_4 \omega'_{it-1} + \beta_5 \omega'_{at-1} + \beta_6 x'_{it-1} - \beta_7 u_t + e'_t \quad (5)$$

where $\omega_{it} = w_{it} - c_t$; $\omega'_{it} = \omega_{it} - x'_{it}$; $\omega'_{at} = \omega_{at} - x'_{it}$

$$\beta_1 = (b_1 - 1);\ \beta_2 = b_2;\ \beta_3 = (b_3 - 1);\ \beta_4 = b_4;\ \beta_5 = b_5;\ \beta_6 = (b_6 - b_4 + b_5);\ \beta_7 = b_7$$

Note that β_1 takes negative values if b_1 is lower than one, while the significance of β_6 indicates departure from linear homogeneity in the steady-state relationship between ω'_{it}, ω'_{at} and x'_{it}. Equation (5), with various measures of outside wages taken as proxies of ω'_{at}, is the baseline specification utilised in empirical work. Tables 27.2, 27.3, and 27.4, supplied in the Appendix, actually refer to restricted versions of this expression.

WAGE RELATIVITIES AND WAGE DETERMINATION: THE RESULTS

The dataset

In the last thirty years, ISTAT (the Italian national statistical institute) has carried out several revisions of its published time series, with the result that homogeneous series of direct ISTAT source are only available from 1970. However, Golinelli and Monterastelli (1990) have reconstructed, for the 1951 – 90 period, annual series of the main national accounting aggregates which are compatible with the currently used definitions. These series have been constantly renewed and are now available for the 1951 – 93 period. It was chosen to base the empirical work on them, as national accounting series are the only ones to allow derivation of sectoral wage, price, output and employment data throughout the whole economy. Furthermore, relying on the Golinelli and Monterastelli dataset seemed preferable to adopting the ISTAT quarterly series beginning in 1970, because of the higher sectoral disaggregation of the annual data (twenty-one industries versus seventeen, and, in particular, eleven manufacturing sectors versus eight).

The national accounting series that have been utilised include sectoral real value added Q_{it}, value added deflator P_{it}, total employment N_{it} and annual labour compensation per employee W_{it}. Output per person, X_{it}, iscalculated as Q_{it}/N_{it}. Unfortunately, there is no readily available series of work-hours to be used for the calculation of output per man-hour and of hourly labour compensation. The private consumption deflator, C_t, was taken as the measure of consumer prices, and I_t, the ratio between direct taxation and nominal GDP, was included in the set of instrumental variables.

The list of the twenty-one industries considered is given below, in the order and with the shortened names that have been utilised in presenting the results. In brackets the proper industry names taken from the Eurostat classification system are also given.

- Agric. (Agricultural, forestry and fishery products)
- Energy (Fuel and power products)
- Metals (Ferrous and non-ferrous ores and metals)
- Non-met. (Non-metallic minerals and mineral products)
- Chem. (Chemical products)
- Engin. (Metal products, machinery and electric goods)
- Vehicles (Transport equipment)
- Food (Food, beverages and tobacco)
- Textiles (Textiles and clothing, leather and footwear)
- Timber (Timber and furniture)
- Paper (Paper and printing products)

- Rubber (Rubber and plastic products)
- Oth. man. (Other manufactured products)
- Constr. (Building and Construction)
- Distrib. (Wholesale and retail trade services)
- Hotels (Lodging and catering services)
- Transport (Transport and communication services)
- Banking (Services of credit and insurance institutions)
- Oth. serv. (Other market services)
- Govt (General government services)
- Non-m.s. (Other non-market services)

The only series utilised in empirical work that does not come from national accounting is the rate of unemployment. While ISTAT begun only in 1959 to carry out regular quarterly labour force surveys, annual labour force surveys had been taken already in 1952, 1954, 1955, 1956, 1957 and 1958. These data have been utilised in order to reconstruct for the 1951 – 8 years employment and unemployment series that could be compatible with those available from 1959 onwards. Moreover, in order to estimate the relations of interest over the 1951 – 93 period, care was also taken of some methodological revisions that have occurred in the 1959 – 93 period. Further details of this work are spelled out in Destefanis (forthcoming) and are available from the author on request.

Some preliminary remarks

In order to explore the properties of the data, the augmented Dickey-Fuller test was applied to variables ω'_{it}, ω'_{at}, x'_{it}, Δc_t, u_t and i_t. The aggregate variables, as well as virtually all the sectoral series turned out to be $I(1)$,[9] meaning that statistical significance of the level terms included in (5) would indicate the existence of a co-integration relationship among these variables. This allows a better understanding of the specification search that has yielded the restricted estimates presented in the Appendix. As a matter of fact, fully fledged estimates of (5) turned out to have poorly determined coefficients in most cases, suggesting the need for a specification search from the general to the particular. This search, like in Pesaran and Lee (1993), was partly guided by *a priori* considerations. First, variable Δc_t was deleted whenever its t-ratio was lower than one in absolute value. The rationale of this procedure is that, unless there is rather strong evidence to the contrary, β_1 is expected to be equal to zero, implying that Δc_t is believed *a priori* to be redundant. In fact, a rather high rate of consumer price inflation has plagued the Italian economy since the early 1970s, attracting workers' attention to the evolution of this variable. Also, most wage settlements included wage escalator clauses even before this period.

Since in this paper the burden of proof rests on the wage leadership hypothesis, the second stage of the search was to delete $\Delta\omega'_{at}$ whenever its t-ratio was lower

than one in absolute value, whereas the third step applied the same procedure to $\Delta x'_{it}$. Differenced variables were deleted before variables in levels, the object being to test thoroughly the existence of co-integration relationships. When considering deletion of terms in levels, the search started with ω'_{at-1} (always because this variable was deemed as irrelevant unless there was decisive evidence to the contrary). The search stopped short of variables ω'_{it-1}, to allow in any case the computation of a steady-state solution, and u_t, given the strong *a priori* interest in the sign and size of the coefficient attached to this variable. Finally note that, for ease of exposition, only restricted versions of (5) are presented in the Appendix, including among the diagnostics a variable deletion test relative to the unrestricted model.

Finally, as Equation (5) was estimated through Instrumental Variables, it seems appropriate to list here the variables which have been used as instruments (beside the constant term):

$$\Delta c_{t-1},\ \Delta c_{t-2},\ \omega'_{it-1},\ \omega'_{it-2},\ \omega'_{at-1},\ x'_{it-1},\ x'_{it-2},\ u_{t-1},\ u_{t-2},\ i_{t-1},\ i_{t-2}$$

As in Pesaran and Lee (1993), only variables likely to belong to a publicly available information set were used as instruments. Adopting the wage leadership hypothesis means that lagged values of ω'_{at} can be reasonably expected to enter the information set of workers and firms. However, only ω'_{at-1} was taken as an instrument, reflecting the judgement that less recent values of the outside wage are already out of this set.

Assessing the evidence

Clearly, the most critical decision in this paper was the choice of the industries to be taken as leading sectors. In this respect, it was thought appropriate to consider, on the one hand, a choice entirely made on *a priori* grounds and, on the other hand, a wholly data-driven approach. As far as the former alternative is concerned, two different sectors were singled out as possible bellwether industries: Engineering and Government. In spite of the lack of formal empirical evidence, the received wisdom in the Italian literature is that Engineering should be considered the most likely candidate for the role of key sector.[10] In recent years, however, this opinion has been increasingly questioned (see for instance Gavosto and Sestito 1991:6). Instead, Government was proposed as the bellwether industry, particularly as far as the events of the last two decades are concerned.

As the present work is based on estimates for the 1951 – 93 period taken as a whole, it cannot possibly hope to give any final evidence on this matter. All the same, each of these two sectors was singled out in turn to assess its performance as a wage leader. The main results of these estimates are shown in Tables 27.2 and 27.3 in the Appendix, and can be summed up as follows.

Generally speaking, the estimates given in Tables 27.2 and 27.3 perform rather well as far as fit and diagnostics are concerned. However, when outside wages were modelled as wages in Engineering (Table 27.2) the fit is higher, most of the times in a quite decisive manner. This points to Engineering being the key sector, a supposition which is largely confirmed by perusal of the estimates of β_2 and β_5. Indeed, in Table 27.2, β_2 is significant (as in the following cases, the conventional 5 per cent level is taken) with the right (positive) sign in eight cases, and has the right sign in two more cases (Government and Other non-market services). For β_5 too there are eight positive and significant coefficients, and four more coefficients with the right sign. In Table 27.3, β_2 is positive and significant in two cases (Other non-market services and Energy) and just positive in four more cases. *In no case* is β_5 positive and significant, while there are four industries in which it has at least the right sign. The upshot is that the present estimates find clear-cut evidence in favour of the wage leadership hypothesis, at least in the secondary sector, if Engineering (rather than Government) is taken as the key sector.

As far as the other variables are concerned, in most cases it was possible to exclude Δc_t from the equations, fulfilling the *a priori* expectations. However, the few significant β_1's seem to have a positive, rather than negative, sign. This can be rationalised on the grounds of a positive functional relationship between the rate of inflation and the proportion of negotiatingworkers, something which is beyond the scope of this work. On the other hand, coefficient β_3 is often negative, implying that rates of change in wages are not linear homogeneous relative to rates of change in productivity.

The coefficient on ω'_{it-1} is very often significant and has virtually always the right sign (on only one occasion, for Government in Table 27.2, is it positive). This indicates that specifications excluding an error correction term are likely to be severely misspecified. Especially in Table 27.2, β_6 is often significant, implying a departure from steady-state linear homogeneity in the relation between real wages and productivity. In no case, however, are these discrepancies very large. As a consequence, the link between productivity and real wages is stronger where the influence of the key sector on wage determination is weaker, both relatively to the contribution of Engineering wages and in absolute terms.

Finally, the estimates reflect the existence of a pervasive influence of the rate of unemployment on wage determination. For instance, in Table 27.2 β_7 is negative and significant in twelve cases, and has the expected sign in six other cases. Only in Agriculture and in Other non-market services is the sign of β_7 positive, whereas even for Government the coefficient is negative (although extremely low in absolute terms).

In order to check the above results and to gain further knowledge on cross-industry wage interactions in the Italian economy, an entirely data-driven approach was also taken to search for possible wage contours. Developing a

suggestion contained in Sargan (1971:60 – 1), the correlation matrix of the rates of change in sectoral real wages was calculated, and for each industry *i* the (potential) leader was chosen as the other industry *j* which gave the highest coefficient of correlation *(ρ)* among rates of change in *Wi/C* and *Wj/C*. These highest *ρ*-industry pairs are shown in Table 27.1 below.

Engineering turns out to be the industry appearing most frequently as a potential leader, yielding further credibility to the results presented in Table 27.2. Also, the highest *ρ*'s are found in manufacturing, confirming that wage-wage interactions should be relatively strong in these industries. Finally, for five industry pairs (Agriculture and Food, Energy and Construction, Engineering and Vehicles, Textiles and Timber, Distribution and Other market services) there is a mutual relationship of highest correlation. Hence these pairs might be contours characterised by a bounded spillover mechanism. In any case, Table 27.4 shows the results obtained by estimating Equation (5) under the assumption that in any industry *i* wages depend on wages from the industry *j* singled out above.

These results by and large confirm those given in Table 27.2, but there are also some interesting differences. Again, fit and diagnostics are rather comforting. In Table 27.4, β_2 is significant and positive in fourteen cases, and has the right sign in one more case (Government). For β_5 there are again eight positive and significant coefficients, and six more coefficients with the right sign. Hence the search for wage contours implies that cross-industry wage interactions become

Table 27.1 Highest-*ρ* industry pairs

Industry I	*Industry j*	*ρ*
Agric.	Food	0.6815
Energy	Constr.	0.6694
Metals	Engin.	0.8703
Non-met.	Engin.	0.7960
Chem.	Engin.	0.7184
Engin.	Vehicles	0.8261
Vehicles	Engin.	0.8261
Food	Agric.	0.6815
Textiles	Timber	0.8195
Timber	Textiles	0.8195
Paper	Chem.	0.7073
Rubber	Textiles	0.7278
Oth. man.	Non-met.	0.6011
Constr.	Energy	0.6694
Distrib.	Oth. serv.	0.7790
Hotels	Constr.	0.5409
Transport	Govt	0.5476
Banking	Oth. serv.	0.6196
Oth. serv.	Distrib.	0.7790
Govt	Transport	0.5476
Non-m.s.	Rubber	0.4264

even more relevant, especially insofar as rates of change are concerned. Pointing the attention to steady-state relationships, no wage-wage interaction appears to exist between Food and Agriculture and in the service sector. Indeed, the only cross-industry wage interactions which were not previously visible (and that subsequently were brought out by the search for wage contours) are those between Textiles and Timber. Also, there is some sign of a wage-wage link going from Vehicles to Engineering.

As for the other variables, it seems worthwhile to mention that the influence of the rate of unemployment on wage determination is less strong than in Table 27.2. Indeed, β_7 is negative and significant in six cases only, and has the expected sign in nine other cases. No other modification of interest occurs relative to the previous results of Table 27.2.

CONCLUDING REMARKS

In the present work the weight of a wage norm and of bargaining power in wage determination has been assessed by estimating the empirical relevance of cross-industry wage comparisons *vis-à-vis* other determinants of wages (such as the rate of consumer price inflation, sectoral labour productivity and the rate of unemployment). A sectoral wage equation was specified for this purpose, and some measures of outside wages included among theregressors. More precisely, in accordance with various versions of the wage leadership hypothesis, outside wages were equated to wages in Engineering, wages in Government and wages chosen following a data-driven procedure proposed in Sargan (1971). This equation was estimated for twenty-one industries of the Italian economy, using annual data from the period 1951 – 93. Finally, it is worth recalling that the specification search carried out on the estimates was designed to put the burden of proof on the wage leadership hypothesis.

The results can be summed up as follows. Sectoral wage equations appear to be rather well specified in general, showing a somewhat pervasive role for sectoral productivity, the rate of unemployment and the rate of consumer price inflation in wage determination. In no case does wage leadership affect wage determination in Agriculture and in Non-metallic minerals and mineral products. In the service sector, the industry most prone to be influenced by wage-wage interactions is Distribution, whereas evidence for other industries of this sector is rather weak (save for rates-of-change effects in Other market services). Wage-wage interactions turn up very significantly in the secondary sector, if Engineering or the industry chosen *à la* Sargan is taken as the key sector. Moreover, there is some evidence of a contour comprising Textiles and Timber, while Engineering comes out strongly as a leader for other sectors in Manufacturing, Construction, Distribution and possibly also Energy.

The choice of the Italian economy as a field of inquiry reflected the desire to test the wage leadership hypothesis in a context where evidence on the matter

had been up to now far from conclusive. Aside from the present paper, it appears that only in Chillemi (1982) has the wage leadership hypothesis been tested within the Italian economy, controlling for the influence of other factors on wages.[11] Chillemi examines the wage leadership hypothesis within the framework proposed in Mehra (1976), utilising quarterly data for manufacturing industries spanning the 1959 – 75 period. A composite leader group, comprising Engineering, Vehicles, Chemicals and Rubber, is selected on the basis of some quantitative criteria (high growth of employment, high growth of productivity, high share of large firms) and of *a priori* judgement. Then, Phillips curves are estimated for all industries and for this group, and the significance of the residuals from the equation for the leader group in explaining residuals from the other equations is assessed. It turns out that evidence in favour of this relationship is at best weak, urging Chillemi to conclude that sectoral wage equations which also include aggregate determinants of wages do not reveal the existence of any wage leadership mechanism. There may be various reasons for the difference between Chillemi's and the present results. The one that seems of more general interest, and that could be worth exploring in future work, is that Chillemi's specification did not include error correction terms, thus failing to account for all the information available in the data.

Of greater interest for future work is the sectoral pattern of the present results, where cross-industry wage comparisons seem to matter mainly in the secondary sector. Clearly, this signals the existence of some kind of segmentation in the labour market, something which is not easily accounted for by a competitive view of the labour market. It has been emphasised above that the existence of a wage-wage spiral cannot be taken *per se* as exhaustive empirical evidence in favour of the post-Keynesian approach to wage inflation. Yet the fact that wage-wage interactions appear to be stronger in industries belonging to the so-called core (or primary) segment of the labour market suggests that the power to enforce a 'fair wage' norm is indeed one of the key factors behind wage determination. More research on this point appears to be warranted, also relying on alternative definitions of labour market segmentation.[12]

As a final reflection, the present exercise has shown how a long-neglected (as far as empirical work is concerned) institutionalist hypothesis has undergone with some success the scrutiny of modern quantitative techniques. It is hoped that this example might elicit further quantitative research that takes seriously the economic role of institutional factors, something that, to quote freely from Geoff Harcourt, would not be social science imperialism but rather choosing the right horse for the course.

ACKNOWLEDGEMENTS

My heartfelt thanks go to Roberto Golinelli, who generously supplied me with his most recent dataset, and to Marco Musella for helpful discussions. The usual disclaimer obviously applies.

NOTES

1 See for instance Harcourt (1974, 1979, 1982).
2 See for instance Wood (1978) or Lavoie (1992).
3 See Harcourt (1974:522), or Lavoie (1992:389).
4 As observed in Dell'Aringa (1983:74) and in Pesaran and Lee (1993:21), the empirical literature on cross-industry wage interactions is far from abundant. In any case, classic references on the subject include Eatwell *et al.* (1974), Mehra (1976) and Sargan (1971).
5 In models where firms are profit maximisers, these claims will be equal to profit-maximising real product wages.
6 See for instance Layard and Nickell (1985). An excellent introductory account to efficiency-wage models can be found in Johnson and Layard (1987). Classic references on formalised union or bargaining models include Nickell (1981) and Oswald (1985).
7 This point is forcefully made in Pesaran and Lee (1993:31 – 4).
8 Taking logs of the rate of unemployment can be seen as a compromise between the linear and hyperbolic specifications which have been rather interchangeably used in the Italian literature on the Phillips curve.
9 The results of these tests, which are not reported here, are available from the author on request.
10 See Dell'Aringa (1983:76) or Gavosto and Sestito (1991:5).
11 Cross-industry wage interactions in the Italian economy are also analysed in detail by Dell'Aringa (1983) and by Gavosto and Sestito (1991). However, these analyses do not allow for the influence of other variables on wages in assessing the existence of wage-wage interactions.
12 See Reich (1984) for a very useful empirical approach to the definition of labour market segmentation.

REFERENCES

Chillemi, O. (1982) 'Sull'ipotesi del gruppo guida della dinamica salariale: un tentative di verifica statistica per i settori industriali in Italia (1965 – 75)', *Note Economiche,* 1982 – 4:147 – 66.

Dell'Aringa, C. (1983) *Prezzi e redditi, frammenti di un modello econometrico disaggregato,* Bologna: II Mulino.

Destefanis, S. (forthcoming) *The Macroeconomic Analysis of Price and Wage Flexibility,* Aldershot: Ashgate.

Dunlop, J.T (1944) *Wage Determination under Trade Unions,* New York: Macmillan.

Eatwell, 1, Llewellyn, J., and Tarling, R. (1974) 'Money wage inflation in industrial countries', *Review of Economic Studies,* 41, 515 – 23.

Eichner, A.S. (1976) *The Megacorp and Oligopoly,* Cambridge: Cambridge University Press.

Gavosto, A. and Sestito, P. (1991) 'Dinamica retributiva e differenziali salariali', *Temi di discussione,* no. 153, Rome: Bank of Italy.

Golinelli, R. and Monterastelli, M. (1990) 'Un metodo per la ricostruzione di serie storiche compatibili con la nuova Contabilità Nazionale', Nota di lavoro no. 9001, Bologna: Prometeia.

Harcourt G.C. (1974) 'The social consequences of inflation', *The Australian Accountant,* 64, 520 – 528.

——(1979) *Economic Issues and the Future of Australia,* DP no. 6, APSA Monograph no. 23, Canberra: APSA.

——(1982) 'Making socialism in your own country', 12th Annual John Curtin Memorial Lecture, Canberra: ANU.

Johnson, G. and Layard, R. (1986) 'The natural rate of unemployment: explanation and policy', in O.Ashenfelter and R.Layard (eds) *Handbook of Labour Economics,* vol. II, Amsterdam: North-Holland, 921 – 99.

Kaldor, N. (1959) 'Economic growth and the problem of inflation: part 2', *Economica,* 26, 287 – 98.

Koenker, R. (1981) 'A note on studentizing a test for heteroskedasticity', *Journal of Econometrics,* 17, 107 – 12.

Lavoie, M. (1992) *Foundations of Post-Keynesian Economic Analysis,* Aldershot: Edward Elgar.

Layard, R. and Nickell, S.J. (1985) 'The causes of British unemployment', *National Institute Economic Review,* no. III, 62 – 85.

Mehra, Y.P. (1976) 'Spillovers in wage determination in US manufacturing industries', *Review of Economics and Statistics,* 58, 300 – 12.

Nickell, S. (1981) 'A bargaining model of the Phillips Curve', DP no. 130, London: LSE Centre for Labour Economics.

Oswald, A.J. (1985) 'The economic theory of trade unions: an introductory survey', *Scandinavian Journal of Economics,* 87, 160 – 93.

Pesaran, M.H. and Lee, K.C. (1993) 'The role of sectoral interactions in wage determination in the UK economy', *Economic Journal,* 103. 21 – 55.

Pesaran, M.H. and Pesaran, B. (1991) *Microfit 3.0: An Interactive Econometric Software Package,* Oxford: Oxford University Press.

Reich, M. (1984) 'Segmented labour: time series hypothesis and evidence', *Cambridge Journal of Economics,* 8, 63 – 81.

Ross, A.M. (1948) *Trade Union Wage Policy,* Berkeley CA: California University Press.

Sargan, J.D. (1964) 'Wages and prices in the United Kingdom: a study in econometric methodology', in P.E.Hart *et al.* (eds) *Econometric Analysis for National Economic Planning,* London: Butterworth, 25 – 63.

——(1971) 'A study of wages and prices in the UK 1949 – 68', in H.G.Johnson and A.R.Nobay (eds) *The Current Inflation,* London: Macmillan, 52 – 71.

Wood, A. (1978) *A Theory of Pay,* Cambridge: Cambridge University Press.

APPENDIX

The coefficients β_i, i=0,…7, refer to specification (5) in the text. The values in brackets below the coefficients are t-ratios. Whenever the diagnostic statistics signalling the presence of serial correlation or heteroscedasticity were significant at the 5 per cent level, the reported t-ratios have been calculated using a consistent estimate for the variance-covariance matrix. R^2_c denotes the coefficient of determination corrected for degrees of freedom; SER is the standard error of the regression.

Diagnostic statistics include: Sargan's general misspecification test, Sarg. $^2(n)$, where n is equal to the number of instruments minus the number of regressors and is given in brackets below the value of the statistic. The LM test for first-order serial correlation, SC $^2(1)$. Ramsey's test of functional form (including squares of the predicted values in the regression), FF $^2(1)$. The Jarque-Bera normality test, J-B $^2(2)$. The heteroscedasticity test proposed in Koenker (1981), HS $^2(1)$. For the reader's convenience, all statistics significant at the 5 per cent level have been underlined.

The statistics denoted by ^{2}(m) relate to a ^{2} test of variable deletion (m is the number of regressors excluded from the structural form and is given in brackets below the value of the statistic).

All estimates have been carried out through *Microfit 3.0,* and further details about the statistics here reported can be found in Pesaran and Pesaran (1991).

Table 27.2a Structural coefficients (outside wages equal to wages in Engin.)

Industry	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Agric.	−0.1482	0.4133	−0.2692	−1.0782	−0.3543	0.0	0.0404	0.0278
	(−2.56)	(2.66)	(−1.24)	(−3.48)	(−3.24)	(–)	(2.22)	(0.57)
Energy	−0.3058	−0.4351	0.0	−1.0944	−0.2383	0.2446	0.0345	−0.1423
	(−1.66)	(−1.26)	(–)	(−7.75)	(−3.75)	(1.54)	(1.45)	(−4.89)
Metals	0.2289	0.0	0.9677	0.0	−0.4782	0.4428	−0.0666	−0.0199
	(2.66)	(–)	(21.91)	(–)	(−3.37)	(3.54)	(−3.23)	(−1.98)
Non-met	−0.5278	0.0	0.0	−0.6931	−0.3774	0.0	0.0441	−0.0268
	(−4.06)	(–)	(–)	(−5.04)	(−3.34)	(–)	(2.99)	(−1.83)
Chem.	−0.0703	0.0	0.5056	−0.4690	−0.3872	0.3817	0.0238	−0.0333
	(−1.18)	(–)	(2.43)	(−1.63)	(−1.89)	(1.75)	(1.48)	(−2.37)
Engin.								
Vehicles	−0.4453	0.0	0.6127	0.0	−0.6111	0.5046	0.1014	−0.0116
	(−3.41)	(–)	(4.44)	(–)	(−3.68)	(3.27)	(3.50)	(−0.97)
Food	−0.4919	0.0	0.4271	0.0	−0.4373	0.4408	0.0881	−0.0648
	(−2.69)	(–)	(2.98)	(–)	(−4.58)	(3.82)	(3.60)	(−4.03)
Textiles	−0.1297	0.0	0.0	−0.7380	−0.1528	−0.1528	−0.0441	−0.0576
	(2.36)	(–)	(–)	(−3.55)	(−1.54)	(−1.07)	(−1.62)	(−2.84)
Timber	−0.2051	0.0	0.0	−0.9732	−0.0991	0.0	0.0	−0.0713
	(−2.33)	(–)	(–)	(−5.75)	(−3.17)	(–)	(–)	(−4.94)
Paper	−0.3066	0.0	0.5294	0.0	−0.5374	0.3687	0.0274	−0.0466
	(−2.34)	(–)	(4.10)	(–)	(−3.21)	(2.75)	(1.77)	(−2.28)
Rubber	−0.3658	0.4206	0.0	−0.8123	−0.7830	0.5210	0.0	−0.0568
	(−5.13)	(3.26)	(–)	(−6.22)	(−4.63)	(3.32)	(–)	(−4.76)
Oth. man.	−0.2268	0.0	0.8494	0.0	−0.3269	0.4131	0.0404	−0.0157
	(−1.27)	(–)	(6.11)	(–)	(−2.79)	(2.77)	(1.36)	(−0.55)
Constr.	−0.5398	0.0	0.0	−0.4691	−0.5349	0.4258	0.0544	−0.0837
	(−3.51)	(–)	(–)	(−1.92)	(−3.48)	(3.43)	(2.23)	(−4.11)
Distrib.	−0.0863	−0.1529	0.4436	−0.6150	20.3583	0.3229	−0.0397	−0.0748
	(−1.03)	(−1.68)	(2.93)	(−2.11)	(21.21)	(2.76)	(−1.27)	(−4.47)
Hotels	−0.3146	0.5865	0.0	−0.5956	−0.4485	0.0	−0.0913	−0.0604
	(−2.69)	(2.73)	(–)	(−1.71)	(−3.58)	(–)	(−3.48)	(−3.44)
Transport	−0.0104	0.0	0.0	−1.0888	−0.1040	0.0	−0.0396	−0.0432
	(−0.09)	(–)	(–)	(−7.41)	(−1.61)	(–)	(−3.04)	(−2.48)
Banking	0.0583	−0.4058	0.0	20.6920	−0.2565	0.0695	−0.0399	−0.0624
	(−0.40)	(−2.08)	(–)	(23.21)	(−1.47)	(1.07)	(−1.65)	(−2.25)
Oth. serv.	−0.2094	0.0	0.5340	−0.2728	−0.5169	0.1809	−0.0409	0.0162
	(−1.43)	(–)	(7.01)	(−1.10)	(−3.13)	(1.39)	(−3.31)	(−0.68)
Govt	−0.0637	0.0	0.2197	0.3171	0.0887	−0.0453	0.0156	−0.0035
	(−1.75)	(–)	(1.53)	(1.83)	(0.79)	(−1.19)	(1.80)	(−0.35)
Non−m. s.	−0.1741	0.0	0.2667	0.0	−0.4237	−0.0249	0.0563	0.0189
	(−1.18)	(–)	(1.59)	(–)	(−3.26)	(−1.06)	(1.01)	(1.34)

Table 27.2b Diagnostic statistics (outside wages equal to wages in Engin.)

Industry	R^2_c	*SER*	*Sarg.*	*SC*	*FF*	*J-B*	*HS*	$\chi^2(m)$
Agric.	0.7998	0.0249	5.7927 (5)	0.4517	2.0668	1.9823	0.9582	0.0202 (1)
Energy	0.7973	0.0345	4.4483 (5)	0.1653	0.0455	2.7809	0.0025	0.1520 (1)
Metals	0.9766	0.0159	2.6633 (6)	0.7021	0.0488	1.6600	1.4350	1.0490 (2)
Non-met.	0.8281	0.0224	4.4086 (7)	0.1954	0.1291	2.0406	0.0238	2.5184 (3)
Chem.	0.8819	0.0241	6.2883 (5)	1.6795	7.7210	0.2190	4.9771	0.2693 (1)
Engin.								
Vehicles	0.8389	0.0211	2.8206 (6)	0.0003	0.0501	1.0526	0.0401	1.1366 (2)
Food	0.6290	0.0227	7.3898 (6)	0.2247	0.1008	1.1659	0.0160	0.1702 (2)
Textiles	0.4888	0.0287	8.0809 (6)	0.0484	0.2598	4.5362	0.0260	0.9139 (2)
Timber	0.6706	0.0274	9.9906 (8)	0.0840	0.4671	7.0636	0.0740	1.3031 (4)
Paper	0.7424	0.0265	4.3891 (6)	1.4252	9.9296	0.2343	1.7905	0.5426 (2)
Rubber	0.7448	0.0277	5.5405 (6)	0.0232	0.0121	0.0090	0.1640	0.8839 (2)
Oth. man.	0.7565	0.0361	0.6600 (6)	0.0491	0.3031	2.1254	0.0002	0.3319 (2)
Constr.	0.5311	0.0258	5.1636 (6)	0.0384	0.2297	1.2113	0.4940	0.3915 (2)
Distrib.	0.7284	0.0222	8.9760 (4)	5.7562	3.1785	2.0598	2.1645	—
Hotels	0.4138	0.0341	3.6692 (6)	0.3645	0.2453	6.6508	0.0094	1.0078 (2)
Transport	0.8002	0.0237	8.1216 (7)	1.0384	0.6607	0.0031	0.4430	1.3780 (3)
Banking	0.6720	0.0341	10.0569 (5)	3.0690	3.2928	8.7982	2.5851	0.0001 (1)
Oth. serv.	0.5961	0.0236	6.6998 (5)	0.3907	2.4370	5.4325	0.7180	0.1119 (1)
Publ. Adm.	0.2227	0.0119	8.1652 (5)	2.8628	0.3652	1.9864	0.7969	0.2017 (1)
Non-m. s.	0.5019	0.0261	5.3170 (6)	3.0780	2.4542	0.0250	13.1826	1.1840 (2)

Table 27.3a Structural coefficients (outside wages equal to wages in Publ. Adm.)

Industry	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Agric.	−0.1495	0.3241	0.0	−0.9456	−0.2922	0.0	0.0333	0.0010
	(−2.98)	(2.29)	(—)	(−5.74)	(−2.89)	(—)	(1.98)	(0.24)
Energy	−0.4562	0.0	0.5613	−0.4418	−0.2930	0.2921	0.0709	−0.0100
	(−3.73)	(—)	(2.15)	(−1.35)	(−2.85)	(1.56)	(1.79)	(−4.70)
Metals	−0.2637	0.0	0.0	−0.8926	−0.0627	0.0	0.0379	−0.0233
	(−1.85)	(—)	(—)	(−9.41)	(−1.66)	(—)	(1.22)	(−1.72)
Non−met.	0.5598	0.0	0.0	−0.6409	−0.4074	0.0	0.0478	−0.0233
	(−4.14)	(—)	(—)	(−4.36)	(−3.44)	(—)	(3.11)	(−1.53)
Chem.	−0.1275	0.0	0.0	−0.9874	−0.0750	0.0	0.0	−0.0374
	(−2.86)	(—)	(—)	(−7.46)	(−1.49)	(—)	(—)	(−2.23)
Engin.	−0.1859	0.1667	0.0	−0.4344	−0.6213	0.0	−0.0649	0.0221
	(−3.08)	(1.72)	(—)	(−2.02)	(−4.05)	(—)	(−3.53)	(1.11)
Vehicles	−0.3975	0.3732	0.0	−0.6687	−0.2681	0.0	0.0456	−0.0067
	(−2.13)	(1.76)	(—)	(−2.47)	(−2.18)	(—)	(1.34)	(−0.29)
Food	−0.4142	0.0	0.0	−1.0055	−0.0745	−0.1660	0.0	−0.1070
	(−2.69)	(—)	(—)	(−2.78)	(−1.42)	(−1.67)	(—)	(−4.88)
Textiles	−0.1952	0.0	0.0	−0.8845	0.1921	0.0	−0.0216	−0.0684
	(−3.11)	(—)	(—)	(−4.49)	(−1.98)	(—)	(−1.15)	(−3.41)
Timber	−0.2073	0.0	0.0	−1.0617	−0.0946	0.0	0.0	−0.0746
	(−3.94)	(—)	(—)	(−5.41)	(−1.62)	(—)	(—)	(−4.85)
Paper	−0.1914	0.0	0.0	−0.7155	−0.1667	0.0	0.0	−0.0351
	(−2.93)	(—)	(—)	(−5.53)	(−2.14)	(—)	(—)	(−2.80)
Rubber	−0.4496	0.2503	0.2784	−0.4361	−0.2790	0.0	0.0264	−0.0533
	(−3.78)	(2.00)	(1.60)	(−1.70)	(−4.22)	(—)	(1.53)	(−3.19)
Oth. man.	−0.5909	0.0	0.3344	−0.4442	−0.2475	0.2299	0.1083	−0.0569
	(−2.17)	(—)	(1.31)	(−1.32)	(−1.99)	(1.72)	(1.75)	(−2.35)
Constr.	−0.1974	0.0	0.2552	−0.6074	−0.1751	0.0	0.0	−0.0325
	(−3.17)	(—)	(1.63)	(−3.02)	(−1.91)	(—)	(—)	(−1.86)
Distrib.	−0.1313	0.0	0.3090	−0.8506	−0.1329	0.1212	0.0	−0.0439
	(−1.70)	(—)	(1.97)	(−3.08)	(−0.98)	(1.54)	(—)	(−2.65)
Hotels	0.3130	0.5875	0.0	−0.6061	−0.4486	0.0	−0.0917	−0.0605
	(−2.65)	(2.66)	(—)	(−1.67)	(−3.52)	(—)	(−3.48)	(−3.43)
Transport	0.0736	0.0	0.0	−0.6061	−0.4486	0.0	−0.0917	−0.0373
	(0.49)	(—)	(—)	(−5.89)	(−0.51)	(—)	(−3.30)	(−1.69)
Banking	−0.0825	−0.3313	0.0	−0.7092	−0.3258	0.1209	−0.0357	−0.0616
	(−0.57)	(−1.15)	(—)	(−3.11)	(−2.45)	(1.21)	(−2.26)	(−3.10)
Oth. serv.	−0.0702	0.0	0.0	−0.6832	−0.2255	0.0	−0.0342	−0.0032
	(−2.05)	(—)	(—)	(−7.79)	(−4.29)	(—)	(−3.90)	(−0.21)
Govt								
Non−m. s.	−0.1475	0.0	0.3487	0.0	−0.3517	0.0	0.0383	0.0072
	(−2.13)	(—)	(3.58)	(—)	(−3.55)	(—)	(1.54)	(0.69)

Table 27.3b Diagnostic statistics (outside wages equal to wages in Gov.)

Industry	R^2_c	*SER*	*Sarg.*	*SC*	*FF*	*J-B*	*HS*	$\chi^2(m)$
Agric.	0.7676	0.0268	4.2655 (6)	1.1051	2.2856	1.6596	0.8171	1.1582 (2)
Energy	0.8099	0.0334	0.4774 (5)	0.8918	0.1347	0.8861	0.2491	0.0159 (1)
Metals	0.8937	0.0308	3.7125 (7)	2.8496	0.4278	16.1304	0.0133	1.3568 (3)
Non-met.	0.8206	0.0229	4.7238 (7)	0.2621	0.0182	2.0807	0.0038	1.6502 (3)
Chem.	0.8013	0.0312	6.2762 (8)	3.6542	5.2842	0.9179	1.5594	0.8675 (4)
Engin.	0.5662	0.0217	6.9779 (6)	3.3916	5.6415	1.2346	2.8968	0.4742 (2)
Vehicles	0.5534	0.0329	4.9757 (6)	1.8508	0.0158	3.8247	2.0343	0.7018 (2)
Food	0.4961	0.0264	11.8492 (7)	2.8120	0.5697	1.1372	0.0377	1.2184 (3)
Textiles	0.4268	0.0304	8.2199 (7)	0.0178	0.0437	6.2623	0.1555	1.3593 (3)
Timber	0.6343	0.0289	8.6766 (8)	0.1135	0.8133	6.1854	0.0	0.5478 (4)
Paper	0.7108	0.0281	4.4111 (8)	3.4121	6.5801	1.1305	0.1974	0.5496 (4)
Rubber	0.7629	0.0267	4.2957 (5)	0.3012	0.9360	3.6030	0.4080	0.8273 (1)
Oth. man.	0.7545	0.0362	1.0195 (5)	1.0195	1.0607	0.8844	2.2805	0.3084 (1)
Constr.	0.3973	0.0292	3.7615 (7)	1.9018	0.4865	0.3534	0.4368	1.1365 (3)
Distrib.	0.6366	0.0257	7.7469 (6)	3.7074	1.2253	1.7243	0.4667	0.2695 (2)
Hotels	0.4128	0.0341	3.8067 (6)	0.4020	0.1983	6.9063	0.0056	0.3370 (2)
Transport	0.6299	0.0323	1.3029 (7)	0.0259	0.6835	0.6703	5.6107	0.9079 (3)
Banking	0.6798	0.0337	8.9319 (5)	1.3580	3.8297	3.7278	4.2571	0.0857 (1)
Oth. serv.	0.5544	0.0248	11.5375 (7)	6.4596	0.0211	10.5334	0.1641	1.1127 (3)
Govt Non-m. s.	0.6217	0.0228	3.6645 (7)	0.5677	2.9653	0.0753	7.7797	2.1056 (3)

Table 27.4a Structural coefficients (outside wages *à la* Sargan)

Industry	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Agric.	−0.1350	0.3587	0.0	−0.8631	−0.3329	0.0	0.0376	0.0291
	(−2.74)	(2.69)	(—)	(−5.10)	(−3.27)	(—)	(2.22)	(0.68)
Energy	−0.1985	−0.3270	0.8847	0.0	−0.2589	0.2136	0.0299	−0.0758
	(−0.94)	(−1.20)	(9.23)	(—)	(−2.63)	(1.34)	(2.38)	(−3.89)
Metals	0.2289	0.0	0.9677	0.0	−0.4782	0.4428	−0.0666	−0.0199
	(2.66)	(—)	(21.91)	(—)	(−3.37)	(3.54)	(−3.23)	(−1.98)
Non−met.	−0.5278	0.0	0.0	−0.6931	−0.3774	0.0	0.0441	−0.0268
	(−4.06)	(—)	(—)	(−5.04)	(−3.34)	(—)	(2.99)	(−1.83)
Chem.	−0.0703	0.0	0.5056	−0.4690	−0.3872	0.3817	0.0238	−0.0333
	(−1.18)	(—)	(2.43)	(−1.63)	(−1.89)	(1.75)	(1.48)	(−2.37)
Engin.	−0.0467	0.1397	0.4228	−0.2718	−0.6166	0.2328	−0.0612	0.0169
	(−0.65)	(1.65)	(2.68)	(−1.66)	(−3.40)	(1.89)	(−2.74)	(1.09)
Vehicles	−0.4453	0.0	0.6127	0.0	−0.6111	0.5046	0.1014	−0.0116
	(−3.41)	(—)	(4.44)	(—)	(−3.68)	(3.27)	(3.50)	(−0.97)
Food	−0.4195	0.0	0.0	−0.9571	−0.1260	0.0	0.0348	−0.0857
	(−1.96)	(—)	(—)	(−2.66)	(−1.31)	(—)	(1.21)	(−4.86)
Textiles	0.1760	0.4014	0.7833	0.0	−0.7363	0.5454	−0.0964	0.0122
	(2.36)	(1.87)	(4.34)	(—)	(−3.11)	(3.20)	(−2.15)	(0.83)
Timber	−0.1806	−0.3213	1.0033	0.0	−0.5771	0.6654	0.0768	−0.0090
	(−2.33)	(−1.21)	(5.76)	(—)	(−3.17)	(2.76)	(1.83)	(−0.63)
Paper	−0.3036	0.0	0.5540	0.0	−0.5808	0.3437	0.0	−0.0283
	(−3.47)	(—)	(5.44)	(—)	(−3.51)	(3.01)	(—)	(−2.17)
Rubber	−0.0499	0.0	0.4698	−0.4012	−0.4762	0.3422	0.0	−0.0224
	(−0.82)	(—)	(2.15)	(−1.63)	(−4.33)	(3.43)	(—)	(−1.96)
Oth. man.	−0.0547	−0.1248	0.8530	0.0	−0.3590	0.4328	0.0	−0.0142
	(−1.00)	(−1.16)	(8.09)	(—)	(−3.29)	(3.15)	(—)	(−0.90)
Constr.	−0.2422	0.3970	0.8726	0.0	−0.5237	0.2698	0.0	0.0708
	(−4.45)	(3.37)	(4.70)	(—)	(−3.88)	(2.44)	(—)	(2.67)
Distrib.	−0.0747	0.0	0.4930	−0.5362	−0.0395	0.0	0.0	−0.0236
	(−1.45)	(—)	(3.23)	(−2.05)	(−0.63)	(—)	(—)	(−1.50)
Hotels	−0.3261	0.6090	0.0	−0.5107	−0.4592	0.0	−0.0918	−0.0593
	(−2.89)	(2.80)	(—)	(−1.67)	(−3.64)	(—)	(−3.43)	(−3.38)
Transport	0.0736	0.0	0.0	−1.3972	−0.0539	0.0	−0.0477	−0.0373
	(0.49)	(—)	(—)	(−5.89)	(−0.51)	(—)	(−3.30)	(−1.69)
Banking	0.1645	−0.4754	0.0	−0.6339	−0.4749	0.2321	−0.0681	−0.0631
	(−1.00)	(−1.16)	(—)	(−3.06)	(−2.22)	(1.67)	(−2.57)	(−2.71)
Oth. serv.	−0.0990	0.0	0.9241	0.0	−0.3693	0.1059	−0.0232	0.0210
	(−1.43)	(—)	(7.01)	(—)	(−3.37)	(1.15)	(−1.65)	(1.43)
Govt	−0.1298	0.0	0.3013	0.2287	−0.1654	0.0824	0.0233	−0.0079
	(−2.47)	(—)	(1.10)	(1.03)	(−0.91)	(1.03)	(1.39)	(−0.65)
Non−m. s.	−0.1290	−0.1919	0.5421	0.0	−0.3903	0.0	0.0505	0.0257
	(−1.52)	(−2.07)	(3.68)	(—)	(−3.55)	(—)	(1.59)	(2.02)

Table 27.4b Diagnostic statistics (outside wages *à la* Sargan)

Industry	R^2_c	*SER*	*Sarg.*	*SC*	*FF*	*J-B*	*HS*	$\chi^2(m)$
Agric.	0.7891	0.0255	3.2008 (6)	0.8009	2.3009	2.3768	0.9917	0.6255 (2)
Energy	0.8761	0.0270	4.3699 (5)	0.8557	0.0530	3.4086	0.0001	0.3060 (1)
Metals	0.9766	0.0159	2.6633 (6)	0.7021	0.0488	1.6600	1.4350	1.0490 (2)
Non-met.	0.8281	0.0224	4.4086 (7)	0.1954	0.1291	2.0406	0.0238	2.5184 (3)
Chem.	0.8819	0.0241	6.2883 (5)	1.6795	7.7210	0.2190	4.9771	0.2693 (1)
Engin.	0.7781	0.0077	6.0160 (4)	0.1105	2.0667	1.2354	0.0539	—
Vehicles	0.8389	0.0211	2.8206 (6)	0.0003	0.0501	1.0526	0.0401	1.1366 (2)
Food	0.4803	0.0268	11.0090 (7)	0.1837	0.3286	0.7049	1.8482	1.5345 (3)
Textiles	0.6585	0.0235	6.6431 (5)	1.9657	0.0621	3.4571	0.5498	0.5090 (1)
Timber	0.7352	0.0246	5.0229 (5)	0.2210	0.0262	2.9934	1.9693	0.7836 (1)
Paper	0.8009	0.0233	2.5263 (7)	0.6028	4.4265	1.3283	0.0147	0.5106 (3)
Rubber	0.8181	0.0234	2.7141 (6)	2.8347	0.0469	0.1506	0.4777	0.1694 (2)
Oth. man.	0.7986	0.0328	1.7099 (6)	0.1859	1.2299	1.6781	0.1282	1.0830 (2)
Constr.	0.5708	0.0246	3.1230 (6)	1.5689	0.9396	0.6736	8.0152	1.1270 (2)
Distrib.	0.7766	0.0211	12.8336 (7)	0.0788	0.1327	0.9290	0.5604	0.4874 (3)
Hotels	0.4039	0.0344	3.7250 (6)	0.4243	0.7283	5.1486	0.0325	1.0401 (2)
Transport	0.6299	0.0323	1.3029 (7)	0.0259	0.6835	0.6703	5.6107	0.9079 (3)
Banking	0.6858	0.0334	8.7074 (5)	3.0164	4.1949	0.5204	5.5345	0.1361 (1)
Oth. serv.	0.7512	0.0185	7.9662 (6)	0.2887	0.1780	0.4593	0.0005	0.0998 (2)
Govt	0.0574	0.0131	11.4509 (5)	0.2384	0.0015	0.4526	0.0421	0.0378 (1)
Non-m. s.	0.5873	0.0238	3.4406 (6)	0.5812	0.0875	1.6177	5.6197	0.5178 (2)

28
DISAGGREGATED IMPORT DEMAND FUNCTIONS FOR THE ITALIAN ECONOMY

Gianluca Cubadda, Stefano Fachin and Francesco Nucci

INTRODUCTION

The beginning of the pure theory of international trade is usually traced back to David Ricardo and his 1817 theory of comparative advantage or, even before, to Adam Smith and his Vent for surplus' argument emphasising potential gains from trade. Empirical investigation of the determinants of international trade flows is a later development, counting among its first contributions J.M.Keynes' pamphlet *The Economic Consequences of Mr Churchill* (1925) on the causes of the deep 1925 recession. This short essay provides a good example of the macroeconomic policy relevance of international trade analyses: Keynes argued that the rise in relative export prices, following the return to the prewar gold-sterling parity, had caused foreign demand of British manufactures to fall despite the high level of world demand.

By now, empirical estimation of import and export functions is a strand of literature representing a considerable body of research whose popularity is partly due to increasing data availability (for a survey, see Goldstein and Khan 1985). Most analyses, however, refer to aggregate rather than to sectoral data (see e.g. Gandolfo and Petit 1983; Giovannetti 1989; Abbot and Seddighi 1996), even though the latter would be more appropriate for a number of reasons. First of all, disaggregated import (export) functions are preferable from the econometric point of view as aggregation bias is directly linked to the level of aggregation. Aggregate elasticities are weighted averages of disaggregated elasticities with weights given by the industries' shares in the aggregate flows, and as such are affected by any change in the weights. In other words, aggregate elasticities may change even if the underlying disaggregated elasticities are constant, certainly an undesirable property (Humphrey 1976). Second, as international trade plays a crucial role inshaping the specialisation patterns of open economies, identifying the sectors for which a comparative advantage of a given country is 'revealed' represents a policy issue of crucial importance, for instance with respect to the Ricardian specialisation versus de-industrialisation debate (see e.g. Barca and

Caselli 1989, and the 'Controversy' section in the January 1996 issue of the *Economic Journal*).

Although both imports and exports contribute to determine the structure of the productive system, in this paper we deal only with the former and focus on the Italian economy, estimating disaggregated long-run import demand functions for manufactures over the period 1977:1 – 1995:4; extensions to exports and other European countries are left to future work. Preliminary to the econometric analysis, we briefly present a simple theoretical framework, based on a standard (see e.g. Goldstein and Khan, 1985) two-stage optimisation problem, whereby sectoral import demands are derived as a function of relative import prices of the corresponding sector and domestic demand for that sector's products. One of the key features of this model is the hypothesis of imperfect substitution between domestic and foreign goods, which is consistent with the large amount of available evidence on intra-industry trade and the failure of the 'law of one price'.

The paper is organised as follows: the simple theoretical framework is briefly recalled in the next section, the empirical results are presented in the following one, while some conclusions are drawn in the last section. Details on the data are given in the Appendix.

THEORETICAL FRAMEWORK

Let us consider an economy composed by n sectors, where domestic demand may be freely addressed to either home-produced or imported goods. These are assumed to be imperfect substitutes so that, consistently with widespread empirical evidence, no Ricardian specialisation results as international trade opens up, even in the long run.

Import demand for consumption and production (intermediate and investment) goods may be seen as the outcome of a two-stage optimisation problem. In the case of consumption, in the first stage each household identifies the utility-maximising demand vector and, in the second, how to allocate its demand for each good between domestically produced and imported goods. Again, if perfect substitution holds, rational behaviour would imply that in the long run products of each country's sector can be either imported or exported but not both (trade specialisation).

In the case of intermediate or investment goods import, demand stems from minimisation of cost functions subject to technology for given output level and input prices (Clements and Theil 1978). As the result is the same, namely that sectoral imports depend on the corresponding domesticdemand and import relative prices, with no loss of generality we concentrate on the case of consumption goods. Under the assumption of a Cobb-Douglas utility function the representative household's problem is the standard one:

$$\text{Max} \qquad u(C_1,..C_n) = A\prod_{i=1}^{n} C_i^{\alpha_i} \qquad (i = 1, 2,.., n) \quad (1)$$

$$\text{s.t.:} \qquad = \sum_{i=1}^{n} p_i C_i = Y \quad (2)$$

where C_i and p_i denote, respectively, real consumption and the price of product i, Y *is* nominal disposable income, and A and α_i $(i=1, 2,\ldots, n)$ *a*re parameters. The utility function is assumed to be linearly homogeneous: $_i a_i=1$.

In the second step sectoral demand (C_i) is assumed to be a CES function defined over domestically produced (CD_i) and imported goods (M_i); each household optimally allocates its product demand among CD_i and M_i, given the solution to the first stage problem:

$$\text{Max} \qquad C_i = \left(\beta_{cd,i} CD_i^{\frac{\theta_i - 1}{\theta_i}} + \beta_{m,i} M_i^{\frac{\theta_i - 1}{\theta_i}} \right)^{\frac{\theta_i}{\theta_i - 1}} \quad (3)$$

$$M_i = C_i \beta_{m,i}^{\theta_i} \left(\frac{p_{m,i}}{p_i} \right)^{-\theta_i} \quad (4)$$

where $p_{cd,i}$ and $p_{m,i}$ denote, respectively, home-produced and import good price indexes;[1] β's and θ_i are parameters, the latter being the elasticity of substitution[2] between CD_i and M_i. The term $\alpha_i Y$ in Equation (4) is domestic demand for product i, that is, the outcome of the previous stage of the problem. Solving (3) and (4) for CD_i and M_i we obtain the following expression for M_i:

$$\text{sub:} \qquad CD_i p_{cd,i} + M_i p_{m,i} = \alpha_i Y \quad (5)$$

where import demand for product i is an increasing function, with unit elasticity, of sectoral domestic demand (C_i) and a decreasing function of import relative price indexes $(p_{m,i}/p_i)$. The term p_i is the price index of sector i's products: it is equal to the sector's unit cost function (uc_i), which can be expressed as (see Uzawa 1962)

$$uc_i = p_i = \left(\beta_{cd,i}^{\theta_i} p_{cd,i}^{1-\theta_i} + \beta_{m,i}^{\theta_i} p_{m,i}^{1-\theta_i} \right)^{\frac{1}{1-\theta_i}} \quad (6)$$

As mentioned above, Equation (5) does not take into account important elements: more precisely, relative prices, albeit with an effect smaller than they

would have in a model with perfect substitution, are the only competitive factors allowed for. Obviously, other forms of international competitiveness do exist and have indeed been emphasised in the literature, such as those hinging on the different qualitative content of products, which may result in a perception of two products of the same category as being not homogeneous (vertical differentiation, see Gabszewicz *et al.* 1981). Another important competitiveness factor hinges on the increasing agents' predilection for differentiated products: the analysis of its impact on international trade has been pioneered by Krugman (1979, horizontal differentiation). By creating profitable opportunities in small market niches, 'love of variety' on the agents' side reduces the relevance of relative import prices and fosters intra-industry trade, i.e. countries tend to simultaneously import and export products of the same category. The model illustrated above is consistent with this pattern of trade; however, the possibility of trade flows in both directions is not adequately micro-founded.

We have previously mentioned that in the model (and in the real world) national price levels for each product do not converge internationally once converted to a given currency.[3] Inefflcacy of international arbitrage to set a 'one world price' both in manufacturing industries and market services is due to a large number of reasons, sometimes sector-specific, only few of which have been spelled out thus far. As an illustrative example, Rogoff (1996) cites both the automobile and the electrical material sectors, for which the likely presence of national standards acts as a soft trade barrier, dampening relative price effects on trade flows. Inefficacy of international arbitrage may thus induce producers to adopt forms of price discrimination in setting prices in different countries (for the 'pricing to market' literature, see e.g. Krugman 1987) and the empirical relevance of this phenomenon (Knetter 1993) inevitably reduces the role of price conditions in assessing competitiveness.

LONG-RUN STRUCTURE OF ITALIAN IMPORTS, 1977 – 95

We now present our estimates of the long-run model implied by the simple theoretical framework discussed in the previous section for the thirteen manufacturing industries corresponding to the two-digit NACE-CLIO classification. Full details on the construction of the data set are given in the Appendix. The period considered (1977:1 – 1995:4) has been chosen in order to exclude the first oil shock and include complete business cycles of the Italian economy. The variables modelled are imports, home demand defined as home production minus net exports, and the relative prices of imports as measured by the ratio of the deflators of imports and home demand. All flows are at constant prices. Plots for the home demand, imports/home demand ratio and relative prices are reported respectively in Figures 28.1, 28.2 and 28.3, while quarterly average growth rates for these variables as well as imports are given in Table 28.1. From

the table and graphs the impressive upward trend followed in all industries but Instruments by the import/demand ratio[4] is evident. Interestingly, the fluctuations following the 1992 exchange-rate crisis, when the Lira left the European exchange-control system and depreciated more than 30 per cent with respect to the US Dollar and the German and French currencies, are largely in line with those previously registered. This is especially noteworthy, as the strong devaluation had a large impact on the relative prices of most industries; the only exceptions are the 'Rubber and plastic' and 'Lumber and furniture' industries, where only a minor upswing may be noticed. Though large, in many cases the 1992 devaluation was not sufficient to offset the cumulated effects on relative import prices of the positive inflation rate differentials and the stable exchange rates of the 1980s. Thus over the entire period relative prices present a mixed picture, with average rates of growth positive in about half of the industries and negative in the other half. Home demand also followed different long-run paths, clearly trending upward in some industries and roughly stationary in others. Common to all sectors is the strong seasonality, a typical feature of Italian economic time-series, mainly due to the concentration of summer holidays in the same period.

Econometric issues

The theoretical background sketched in the previous section provides predictions on sectoral steady states among imports, home demand, and relative prices. Hence we resorted to multivariate cointegration analysis, and specifically to the maximum likelihood method proposed by Johansen (1988). As is well known, this method assumes that the variables of

$$A^{*}(L)(1 - L)x_t = \mu_t - Ax_{t-1} + \varepsilon_t$$

interest admit the following error-correction model (henceforth, ECM):

$$A^{*}(L)(1-L)x_t = Ax_{t-1} + \varepsilon_t \qquad (7)$$

where in our case x_t is a trivariate time series including logarithms of sectoral imports (m_t), domestic demand (d_t), and relative prices (p_t), μ_t is a vector of deterministic elements including an intercept, a linear trend and three seasonal dummies, ε_t is a Gaussian white-noise process, and finally $A^{*}(L)$ is a polynomial matrix of order p in the lag operator L. The rank of the matrix A, i.e. the cointegration rank, is the number of equilibrium relations[5] existing among the variables of interest. Now, our theoretical model predicts the existence of a single steady-state relation; hence, when $r>1$ the cointegration matrix α is not identified on economic grounds.[6] When this case occurred in our empirical

Table 2.1 The Malthus system

Industries	*Imports*	*Home demand*	*Imports/ home demand*	*Relative import prices*
Instruments	2.35	2.43	−0.08	0.49
Machinery	1.38	−0.21	1.59	−0.11
Primary metal	1.00	0.44	0.66	0.17
Chemicals	1.41	0.63	0.82	0.23
Transportation equipt	1.53	0.58	0.98	0.54
Electric equipt	2.03	0.72	1.31	0.26
Rubber and plastic	1.85	−0.02	1.87	−0.38
Textiles, apparel, leather and shoes	1.52	−0.41	1.93	−0.31
Food, beverages and tobacco	0.98	0.49	0.49	0.01
Paper and printing	1.43	0.86	0.57	−0.06
Stone	1.22	0.11	1.11	0.35
Lumber and furniture	1.07	0.24	0.83	−0.27
Fabricated metal	1.72	−0.47	2.19	−0.05

Note: Imports and home demand in real terms. Detailed data definitions and sources are given in the Appendix.

results, we identified the cointegration vectors by using zero-restrictions in order to let the estimated coefficients be interpretable as partial long-run elasticities.

Something more should be said about the deterministic kernel μ_t appearing on the right-hand side of Equation (7). Since this kernel includes linear time-trends, in our analysis we allow for 'stochastic cointegration' (Perron and Campbell 1993). More precisely, evidence favourable to a cointegration rank equal to r suggests that there are r linear combinations of the considered variables which appear to be stationary around a deterministic path. This issue is particularly relevant for our analysis, in view of the strong upward trendingbehaviour of the sectoral import/home demand ratio. Since we introduce the linear trends without restricting them into the error-correction $\alpha' x_p$ we allow for a polynomial trend of order at most two in the process x_t and for a polynomial trend of order at most one in the error-correction term (Johansen 1994).

Moreover, it should be noted that our set-up allows for the possibility oftrend-stationary variables. In fact, in this particular case there exists a 'trivial' cointegrating vector, i.e. such that only the element relative to the I(0) variable is not null. The existence of trivial cointegration vectors can be practically detected by means of likelihood-ratio tests for linear restrictionson the cointegration space (Johansen 1988). Hence no preliminary unit root tests, with the associated risk of size distortion of the cointegration tests induced by the pre-testing, are needed.

We emphasised above the strongly seasonal profile of some of the series modelled: in order to make the Johansen method invariant to the

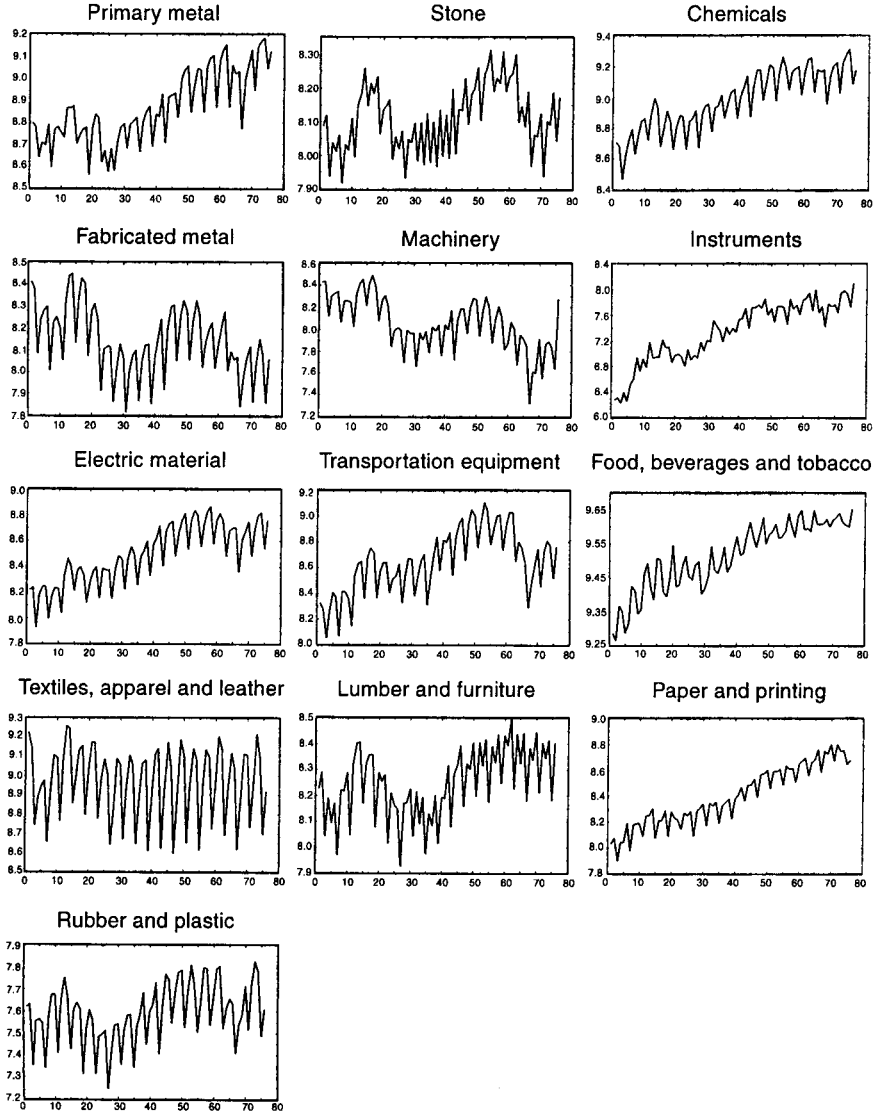

Figure 28.1 Home demand in Italian manufacturing industry, 1977 – 95

seasonalcharacteristics of the process x_t we will include a set of seasonal dummies in the deterministic kernel. In fact, the presence of the seasonal dummies removes a deterministic seasonal pattern without affecting the distribution of the Johansen tests (Johansen 1991). Moreover, these tests are invariant to the presence of stochastic seasonality if the ECM order includes the seasonal lags (Lee 1992). Consequently, we selected the augmentation of the ECM by starting with an initial order equal to eight, thus ensuring that the

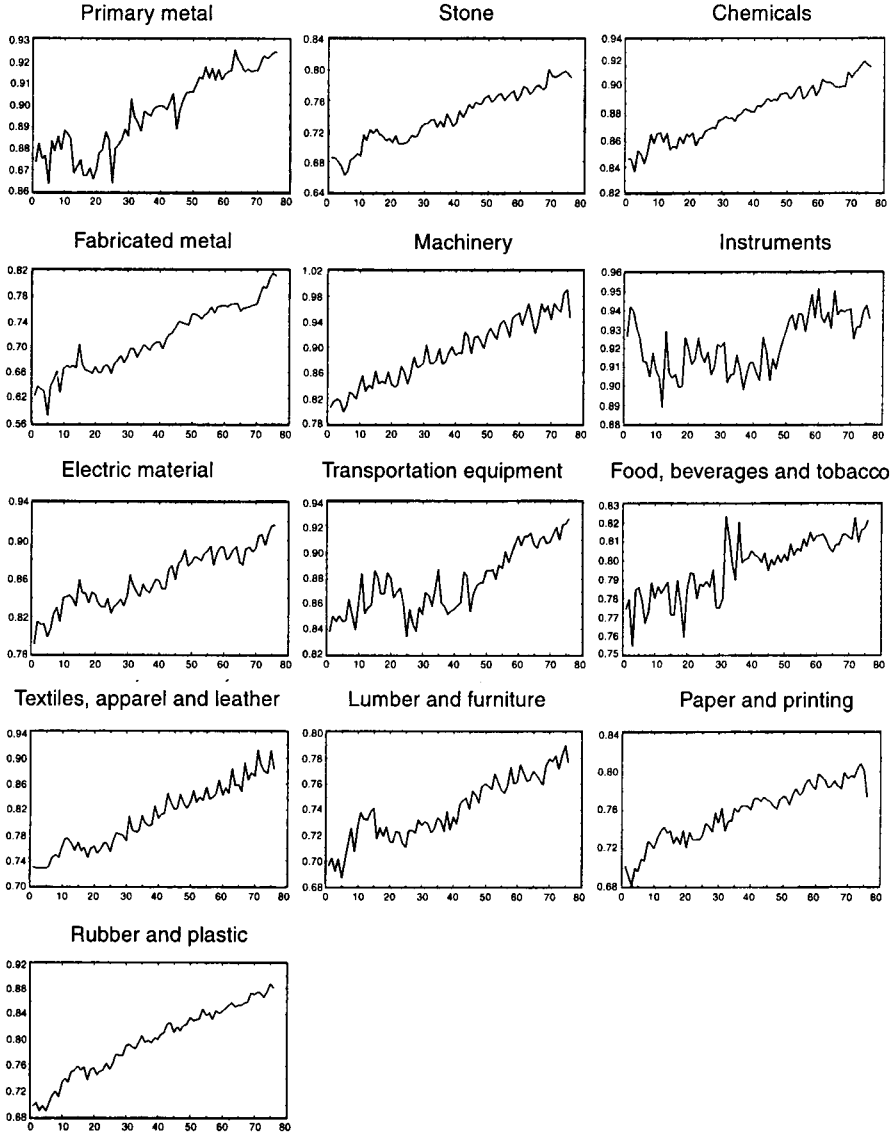

Figure 28.2 Imports/home demand in Italian manufacturing industry, 1977 – 95

variables' seasonal lags would be selected if they were found to be statistically significant.

Finally, it is worthwhile to remark that our approach includes, as particular cases, some data-generation processes of variables x_t which were adopted in previous empirical literature. In fact, Barca and Caselli (1989) implicitly assume a full-rank impact matrix A (i.e. r=3), whereas Committeri *et al.* (1995) allow for a reduced-rank of this matrix in their analysis, though limited to the case r<2.

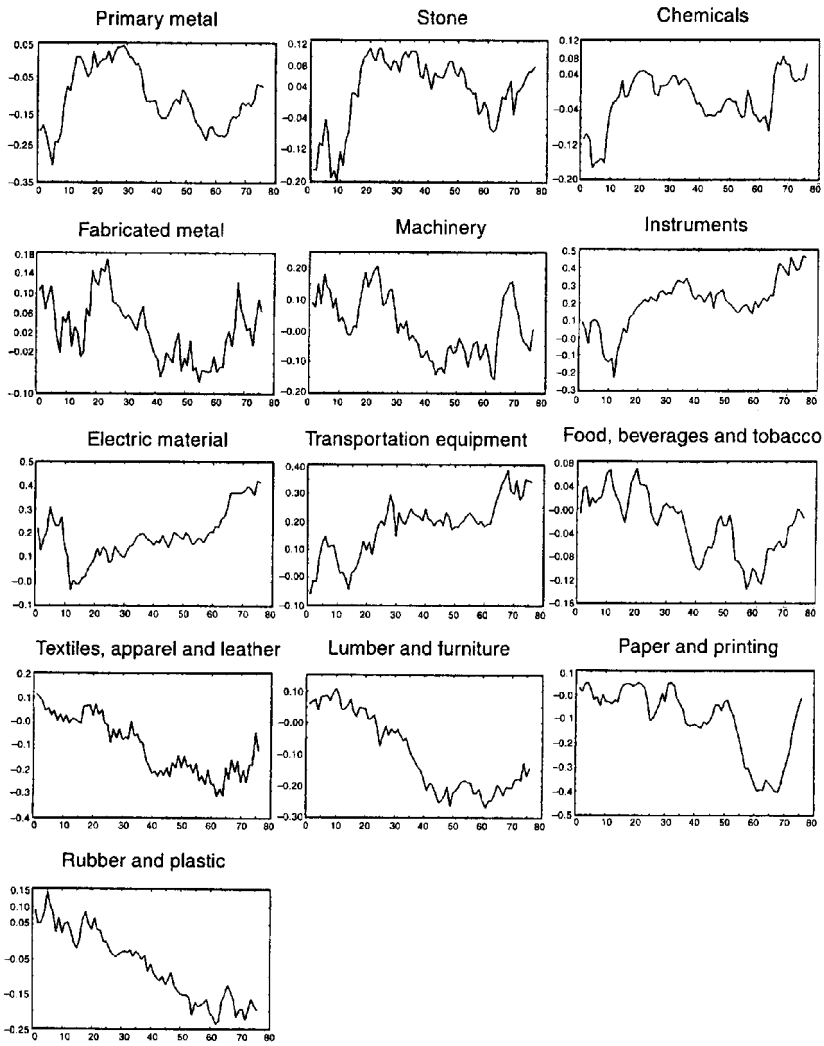

Figure 28.3 Relative import prices in Italian manufacturing industry, 1977 – 95

Empirical results

Let us now turn to the empirical results. Since full details of the VAR estimation for all the thirteen industries cannot be reported here, we merely point out that in all cases we could select a satisfactory specification presenting no misspecification signs.[7] We then tested for the existence of stochastic long-run equilibrium relationships and estimated their coefficients. The results for the cointegration tests and vectors are summarised in Tables 28.2 and 28.3

respectively. The former include both the maximum eigenvalue and the trace tests, as in some cases borderline results advised checking both tests and favouring the clearer-cut result (for instance, as in the case of the Instruments industry, where the first maximum eigenvalue statistic is only marginally not significant but the first trace statistic is strongly significant).

In the majority of cases (eight out of thirteen) all variables turned out to be non-stationary and some non-trivial cointegrating relationships could be identified; in three industries, namely 'Chemicals', 'Food, beverages and tobacco' and 'Rubber and plastic', the hypothesis that all variables are stationary around linear deterministic trends could not be rejected; finally, in two cases, 'Electrical Equipment' and 'Textiles, apparel, leather and shoes', no stable long-run equilibrium relationships seem to exist. A closer examination of the latter industry (due to its importance: about 20 per cent of the gross product of the Italian manufacturing industry in the mid-1980s) reveals an overall rather singular picture. Imports grew constantly despite a stagnating home demand; the almost continuous fall of relative prices may suggest a case of deindustrialisation with imports crowding out home production, but this would imply a bivariate equilibrium relationship between relative prices and imports, which has not been found. A possibleexplanation of this puzzle is given by the role of exports, which also grew significantly over the period from a level only slightly higher than 30 per cent of home demand in the early eighties to 50 per cent at the end of the period. Growing imports *and* exports with stationary home demand may suggest increasing delocalisation of home production, both for internal and external markets, in low-wage countries in the Far East or, especially after 1989, in Eastern Europe. If that is the case, the equilibrium relationship will involve import, exports, relative prices and home demand, and the lack of a long-run equilibrium in our system would be due to an omitted variable.

Leaving further examination of this issue to future work, let us turn to the industries where we did find long-run equilibria. In five of them (namely 'Machinery', Taper and printing', 'Lumber and furniture', 'Fabricated metal' and 'Stone') the cointegration vector is unique and Likelihood Ratio tests did not reject the hypothesis that the coefficient of home demand is 1 and that of relative prices zero.[8] All the significance levels of the tests (reported in Table 28.3) but one exceed 10 per cent, and two are in fact well over 50 per cent. In view of the widespread suspect that asymptotic tests on cointegrating coefficients suffer from excess rejection (see *inter alia* Fachin 1996), and thus that the estimated significance levels are likely to be lower bounds for the true significance levels, this evidence lends very strong support to our theoretical framework in the restricted version, with zero elasticity of substitution between home-produced goods and imports. This is clearly at odds with the traditional view of relative prices as the key competitive factor in international trade and consistent with the evidence on the relevance of intra-industry trade. In other words, in these five industries quality (both in the sense of 'quality of construction' and variety)

appears to be the prevailing long-run competitive factor. While the prevalence of the quality factor is not particularly surprising in the case of highly complex manufactures such as 'Machinery', it is less obvious in the case of apparently undifferentiated goods such as 'Stone' and 'Lumber', for which there is apparently less scope for both vertical and horizontal differentiation. Interestingly enough, these goods are only marginally traded (on average, total international trade in both these industries over the period has been less than 25 per cent of home production, the smallest along with the Food industry), the likely reason being a high transport costs/value ratio due to a low value/weight ratio. In other terms, transport costs of these goods are probably so high with respect to other costs as to make imports non-competitive unless they are qualitatively different (either vertically or horizontally) from home supply, in which case relative prices will matter less, or even not at all. To put it very simply: bricks are not internationally traded, but hand-made or designer tiles are. Indeed, almost all high-quality tiles purchased in the world are made in a small area in the north of Italy (a typical example of a Marshallian 'industrial district').

For Transportation equipment' and 'Primary metal', relative prices do matter while home demand matters less. For the former the effect of relative prices is particularly strong and that of home demand particularly small, suggesting that cars and other transportation equipment tend to be perceived both by consumer and firms as largely undifferentiated products, with home-produced and imported goods basically competing for the same demand. Of course, it may well be the case that this particularly high elasticity is simply due to this industry being more homogeneous than others. The only way to investigate this point would be to repeat the exercise with a more disaggregated dataset. Finally, in the case of 'Primary metal', the relevance of price competitiveness was completely expected in view of the recent history of this industry: large cuts in employment and capital scrapping have taken place in all European countries in the last decade because of the strong competition from the newly industrialised countries, e.g. South Korea.

Two cointegration vectors have been found only in one industry, namely 'Instruments'. The two vectors suggest that, first, there is a long-run link between imports and relative prices, as the linear combination $m-0.6p$ is stationary around a deterministic trend; and second, home demand is also stationary around a deterministic trend. Combining these two pieces of evidence we can conclude that the ratio between imports less the nonstationary part due to the relative price movements and home demand is also stationary (in logs) around a trend.

Before turning to the overall conclusions of our study, we should explain why we did not investigate any further—for instance by estimating regressions of imports on home demand, relative prices and a trend—the three industries where all variables have been found to be stationary around deter ministic trends. The reason is simply that this type of regression has no interest for our objective, as the driving force of the long-run growth of imports in these cases is given by the

Table 28.2 Cointegration tests

H_o: *number of vectors =*	*Cointegration test*	*Instruments*	*Machinery*	*Primary metal*	*Chemicals*	*Transportation equipment*	*Electric material*	*Rubber and plastic*
0	ξ	22.52	28.18*	31.43**	45.58**	23.67	18.1	29.08**
	η	44.62**	42.07*	53.11**	67.07**	35.28*	28.56	52.04**
1	ξ	18.53*	9.37	13.31	17.33*	7.86	–	17.92*
	η	22.10*	13.90	21.68*	21.49*	11.61	–	22.95*
2	ξ	3.57	–	–	4.16*	–	–	5.04*
	η	3.57	–	–	4.16*	–	–	5.04*
Accepted hypothesis:	*Number of vectors =*	2	1	1	3	1	0	3

H_o: *number of vectors =*	*Cointegration test*	*Textiles, apparel, shoes and leather*	*Food, beverages and tobacco*	*Paper and printing*	*Stone*	*Lumber and furniture*	*Fabricated metal*
0	ξ	21.94	34.27**	21.88	32.68**	24.17*	25.88*
	η	33.81	60.5**	39.47*	137.63*	34.14	36.68*
1	ξ	–	20.52*	12.05	10.51	8.90	8.39
	η	–	26.23**	17.59	11.85	9.97	10.8
2	ξ	–	5.71*	–	–	–	–
	η	–	5.71*	–	–	–	–
Accepted hypothesis:	*Number of vectors =*	0	3	1	1	1	1

Notes

Estimation period 1977:1 – 1995:4 including the eight initial observations. Imports and home demand at constant 1980 prices. Industries are listed in decreasing order of imports/home demand ratio. All VARs include up to eight lags, seasonal dummies and unconstrained trends and constants. ξ is the maximum eigenvalue statistic and η the trace statistic of Johansen (1988). *: significant at 5%; **: significant at 1%.

Table 28.3 Cointegrating vectors

	Imports	*Home demand*	*Relative prices*	*Prob (restr)*	*Trend*
Instruments	1	0	0		0.14%
	0	−1	0	0.63	1.43%
Machinery	1	1	0	0.20	1.82%
Primary metal	1	−0.593	0.958	–	0.76%
Transportation equip't	1	−0.274	2.452	–	2.12%
Paper and printing	1	−1	0	0.10	1.69%
Stone	1	−1	0	0.26	1.24%
Lumber and furniture	1	−1	0	0.12	0.77%
Fabricated metal	1	−1	0	0.03	1.91%

Notes
Prob(restr) is the asymptotic estimate of the significance level of an LR test of the restrictions imposed on some vectors.
The coefficients of the deterministic trend, i.e. the quarterly rates of growth of imports for given home demand and relative prices, are reported in the last column.

linear trend. The 'long-run elasticities', that can be computed as the sum of the coefficients of a given independent variable divided by the complement to unity of the sum of the coefficients of the lagged values of the dependent variable, simply measure the ultimate effect that a unit impulse of the former has on the latter: a very different thing from the long-run elasticities estimated in the cointegration framework, which measure the long-run equilibrium relationships between the different variables.

CONCLUSIONS

The results obtained seem to be interesting in many respects. In nearly half of the industries examined, including goods as diverse as 'Machinery' and 'Paper', relative prices do not appear to have played any role in shaping the long-run pattern of imports. This is consistent with the evidence on intra-industry trade, and implies that in these industries competition has shifted from prices to quality, both in the sense of quality of construction and variety of supply. How much this result is merely a consequence of the fairly high level of aggregation considered in this paper is of course an open, and not new, question (it is discussed *inter alia* in Vona 1991) that could partly be answered by analyses based on highly disaggregated datasets. However, this does not diminish the novelty of our results: existing studies at the same level of disaggregation usually reported significant effects of relative import prices.

The simple analytical scheme of a two-stage allocation process, with relative prices ruling the second-stage choice between home-produced and imported goods fully holds only in three industries; the role of relative prices is highest in

'Transportation equipment', an interesting aspect for the future prospects of the European car industry. Finally, there are reasons to suspect that the delocalisation of production of clothes and shoes to low-wage countries has now become so important that neglecting it prevents one finding any stable long-run equilibrium relationship for the textile industry, one of the most important in the Italian economy. This points to the way in which the analysis of international trade flows should be carried out: the simultaneous empirical modelling of both import and export flows.

ACKNOWLEDGEMENTS

We would like to thank G.Carbonaro, M.Committeri, G.Venanzoni and all the participants to the Turin 1997 Conference of the *Società Italiana diStatistica* for helpful comments and discussions; the usual disclaimers apply. Financial support from CNR and MURST is gratefully acknowledged by the second author.

NOTES

1 Note that the dynamics of import price indices includes change in both foreign currency prices and exchange rates. Currency appreciation and depreciation are thus considered in the analysis, although not separately from changes in foreign currency prices. This implies that we will not be able to examine the 'pass-through', i.e. how (and how much) exchange-rate changes are passed to both domestic and import price levels.

2 The elasticity of substitution between products of different sectors is equal to one. The property that two-step utility functions have to satisfy is strong separability, namely that expenditure allocation for a product depends only on relative import prices of that sector (see McFadden 1963; Sato 1967).

3 As the model presented is a partial equilibrium framework, in principle it would not legitimate conclusions of this type; however, it is easy to extend the model to a general equilibrium context and verify that the analysis presented thus far is indeed compatible with non-convergence in price levels (see Bhagwati and Srinivasan 1992).

4 In this study we are not interested in assessing how much of the growth in import penetration is due to market globalisation and how much to a deindustrialisation process with a rise in imports only. On the deindustrialisation of the Italian economy, cf. *inter alia* Fachin and Vichi (1994).

5 On the links between the notion of cointegration and the various concepts of economic equilibria, see e.g. Pesaran (1997).

6 On the issue of economic identification of multiple cointegration vectors, see e.g. Wickens (1995).

7 The dynamic structure of the VARs has been separately selected for each equation on the basis of the usual joint significance tests for all variables at each lag, checking autocorrelation by means of LM tests, heteroskedasticity by means of

both ARCH and White (1980) tests, and normality by means of the Doornik and Hansen(1994) test.

8 In the studies by Barca and Caselli (1989) and Commiterri *et al.* (1995) mentioned above, the elasticities of imports with respect to relative prices in these industries have generally been estimated to be different from zero. However, the former applied the then-standard techniques now known to be inadequate in the case of non-stationary variables, and the latter the Engle-Granger (1987) two-step cointegration method in which significance tests are not available.

REFERENCES

Abbot, A.J. and Seddighi, H.R. (1996) 'Aggregate imports and expenditure components in the UK: an empirical analysis', *Applied Economics,* 28:1119 – 1125.

Bhagwati, J. and Srinivasan, S.N. (1992), *Lectures on International Trade,* Harvard MA: MIT Press.

Barca, F. and Caselli, P. (1989) 'Competitività internazionale e ristrutturazione della industria italiana negli anni '80', *Politico, economica,* 2, 231 – 72.

Clements, K.W. and Theil, H. (1978) 'A simple method of estimating price elasticities in international trade', *Economics letters,* 1, 361 – 4.

Committeri, M., Ferrando, A. and Pozzolo, A. (1995) 'II modello di specializzazione internazionale dell'industria italiana dagli anni ottanta alia svalutazione della lira', mimeo, Rome: Bank of Italy.

Doornik, J.A. and Hansen, H. (1994) 'A practical test for multivariate normality', mimeo, Nuffield College, Oxford.

Engle, R.F. and Granger, C.W.J. (1987) 'Co-integration and error-correction representation: estimation and testing', *Econometrica,* 55, 251 – 76.

Fachin, S. (1996) 'Asymptotic and bootstrap tests of long-run relationships in cointegrated systems', paper presented at the EC(2) conference on 'Simulation Methods in Econometrics', Florence.

Fachin, S. and Vichi, M. (1994) 'Deindustrializzazione, specializzazione o ristrutturazione? Una analisi multiway in matrici fattoriali dell'industria manifatturiera italiana dal 1971 al 1983', *Politica Economica,* 10, 373 – 404.

Gabszewicz, J., Shaked, A., Sutton, J. and Thisse, J.F. (1981) 'International trade in differentiated products', *International Economic Review,* 22, 527 – 34.

Gandolfo, G. and Petit, M.L. (1983) 'The functional form of the aggregate import demand equation: Italy, 1960 – 80', *Economics Letters,* 11, 145 – 8.

Giovannetti, G. (1989), 'Aggregate imports and expenditure components in Italy: an econometric analysis', *Applied Economics,* 21, 957 – 71.

Goldstein, M. and Khan, M.S. (1985) 'Income and price effects in foreign trade', in R.W.Jones and B.Kenen (eds) *Handbook of International Economics,* vol. II, Amsterdam: Elsevier.

Humphrey, D.H. (1976) 'Disaggregated import function for the UK, West Germany and France', *Oxford Bulletin of Economics and Statistics,* 4, 281 – 98.

Johansen, S. (1988) 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamic and Control,* 12, 231 – 54.

——(1991) 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica,* 59, 1551 – 80.

——(1994) 'The role of a constant and linear terms in cointegration analysis of nonstationary variables', *Econometric Reviews,* 13, 205 – 29.

Keynes, J.M (1925) 'The economic consequences of Mr Churchill', London: Hogarth Press. Reprinted in *The Collected Writings of John Maynard Keynes,* vol. IX, 207 – 30, London: Macmillan.

Knetter, M.M. (1993) 'International comparisons of price-to-market behavior', *American Economic Review,* 83, 473 – 86.

Krugman, P. (1979) 'Increasing returns, monopolistic competition and international trade', *Journal of International Economics,* 9, 469 – 79.

——(1987) 'Pricing to market when the exchange rate changes', in S.W.Arndt and J. D.Richardson (eds) *Real-Financial Linkages among open economies,* Cambridge MA: MIT Press, 49 – 70.

Lee, H.S. (1992) 'Maximum likelihood inference on cointegration and seasonal cointegration', *Journal of Econometrics,* 54, 1 – 47.

McFadden, D. (1963) 'Constant elasticity of substitution production functions', *Review of Economic Studies,* 30, 73 – 83.

Perron, P. and Campbell, J.Y. (1991) 'A note on Johansen's cointegration procedure when trends are present', *Empirical Economics,* 18, 777 – 89.

Pesaran, M.H. (1997) 'The role of economic theory in modelling the long run', *Economic Journal,* 107, 178 – 91

Rogoff, K. (1996) 'The purchasing power parity puzzle', *Journal of Economic Literature,* 34, 647 – 68.

Sato, K. (1967) 'A two-level constant-elasticity-of-substitution production function', *Review of Economic Studies,* 34, 201 – 18.

Uzawa, H. (1962) 'Production functions with constant elasticities of substitution', *Review of Economic Studies,* 29, 291 – 9.

Vona, S. (1991) 'Intra-industry trade: a statistical artefact or a real phenomenon?', *Banca Nazionale del Lavoro Quarterly Review,* 43, 383 – 412.

Wickens, M.R. (1996) 'Interpreting co-integration vectors and common stochastic trends, *Journal of Econometrics,* 74, 1637 – 48.

White, H. (1980) 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica,* 48, 817 – 38.

APPENDIX: DATA SOURCES AND DEFINITIONS

Data cover all the thirteen manufacturing industries corresponding to the two-digit NACE-CLIO classification for the period 1977:1 – 1995:4.

Imports (cif): values at constant (1980) prices. Source: unpublished National Institute of Statistics (ISTAT) data.

Exports (fob): values at constant (1980) prices. Source: unpublished ISTAT data.

Home demand: Defined as the value of output minus the difference between exports and imports. For each industry the quarterly series of output at constant (1980) prices has been constructed re-scaling the industrial production index (source: ISTAT, *Numeri Indict delta Produzione Industrial: Base 1990=100,* Rome, 1996) with a scale factor given by the value of output as given in the 1980 input-output table (ISTAT, *Tavola Intersettoriale dell'Economia Italiana: Anno 1980,* Rome, 1984).

Relative import price index: Given by the ratio of imports deflator and home demand deflator. Home demand at current prices is obtained as the value of output minus the difference between exports and imports, all at current prices (source: ISTAT, unpublished data). The quarterly series of output at current prices has been constructed multiplying the corresponding series of output at constant prices by the relevant manufacturing output price index (source: ISTAT, *Numeri Indict del Prezzi alia Produzione del Prodotti Industriali: Base 1980=100,* Rome, 1990, and subsequent updates).

INDEX