

# **ECONOMIC FOUNDATIONS OF STRATEGY**

By

**JOSEPH T. MAHONEY**  
University of Illinois at Urbana-Champaign  
College of Business  
Department of Business Administration  
339 Wohlers Hall  
1206 South Sixth Street  
Champaign, IL 61820

## Table of Contents:

Editor's Introduction	pp.	004 - 005
Foreword	pp.	006 - 008
Preface	p.	009
Overview	pp.	010 - 011
Chapter 1 Behavioral Theory of the Firm	pp.	012 - 014
• Barnard (1938)	pp.	015 - 019
• Simon (1947)	pp.	020 - 041
• March and Simon (1958)	pp.	042 - 057
• Cyert and March (1963)	pp.	058 - 065
Application: Cuban Missile Crisis	pp.	066 - 067
• Simon (1982)	pp.	068 - 085
Chapter 2 Transaction Costs Theory	pp.	086 - 088
• Arrow (1974)	pp.	089 - 105
• Coase (1988)	pp.	106 - 109
Application: The Fable of the Bees	p.	110
• Williamson (1975)	pp.	111 - 131
• Williamson (1985)	pp.	132 - 144
Application: Aerospace Industry	p.	145
• Williamson (1996)	pp.	146 - 154
Appendix: Mutual (Sunk-cost Commitment)	pp.	155 - 157
Chapter 3 Property Rights Theory	pp.	158 - 160
• Libecap (1989)	pp.	161 - 171
• North (1990)	pp.	172 - 179
• Barzel (1989)	pp.	180 - 182
• Eggertsson (1990)	pp.	183 - 186
• Hart (1995)	pp.	187 - 197
Application: Automobile Industry	pp.	198 - 199

Chapter 4	Agency Theory	pp.	200 – 205
	• Berle and Means (1932)	pp.	206 - 208
	• Pratt and Zeckhauser (1985)	pp.	209 - 214
	• Arrow (1985)	pp.	215 - 221
	Application: Oil and Gas Industry	pp.	222 - 223
	• Appendix: Levinthal (1988)	pp.	224 - 233
	• Appendix: Jensen and Meckling (1976)	pp.	234 - 236
Chapter 5	Resource-based/Dynamic Capabilities	pp.	237 - 239
	• Penrose (1959)	pp.	240 - 252
	• Chandler (1990)	pp.	253 - 259
	• Itami and Roehl (1987)	pp.	260 - 262
	• Nelson and Winter (1982)	pp.	263 – 276
	Applications: RBV and Dynamic Capabilities	pp.	277 - 286
	• Trigeorgis (1997)	pp.	287 - 295
	Applications: Real Options	pp.	296 - 298
Chapter 6	The Theoretic Building Blocks of Organizational Economics	pp.	299 - 306
References:		pp.	307 – 329
Biographical Sketch:		p.	330

Editor's Introduction to *Economic Foundations of Strategy*

One sign of increasing maturity in the strategy field is an increasing attention to cumulative, rather than individual theory development. Our research journals are requesting clearer theoretical foundations and demanding more significant theoretical contributions. We are systematizing and taking more seriously our sources in economics, sociology, psychology, and other fields. As a result, we are clarifying the nature of the academic discipline. We are better prepared to contribute to practice. And, I believe we are in a position to begin giving something back to the base disciplines from which we draw.

The FOS series contributes to this important development by reviewing the foundations of the strategy field in economics, the behavioral sciences and psychology. One objective is to specify the most important sources in each tradition for the use of students, and those more distant from the field, having an interest in our subject matter. But a more pressing purpose is to provide the tools for active involvement in theory development. It is our hope, and belief, that the sophistication and standards of the best work from the past will inspire and support the best work for the future.

In this volume, Joe Mahoney identifies five theories of particular importance to the strategy field, and outlines the contributions of particularly influential works in each area. As Oliver Williamson notes in his Foreword, these authors provide alternative theories of the firm. To some extent they are complementary accounts, but they also propose alternatives with somewhat different purposes in mind. The book can be used to understand the key issues for strategy raised by theorists taking an economics perspective and hopefully followed by an attempt to further advance answers to five important questions:

1. How can organizations operate efficiently? — the basic question of the Behavioral Theory of the Firm.
2. How can firms minimize costs? — the basic question of Transaction Costs Theory.
3. How can firms create and allocate wealth? — the basic question of Property Rights Theory.
4. How can firms align individual self-interest? — the basic question of Agency Theory.
5. How can resources be acquired, developed and deployed to improve the likelihood of survival and profitable growth? —the basic questions of Resource Based Theory, and more recently Dynamic Capabilities and Real Options Theories of the Firm.

These are big questions, still only partially answered, and the theory of the firm is of contemporary interest as an organized way to answer them. The questions asked also articulate some of the most basic concerns of managers. As Joe notes in his concluding chapter, the works cited here thus provide important building blocks for new theoretical developments in the field. Thus, our 'Foundations' title applies.

Anne S. Huff

October 2003

## Foreword

An Economics of Organization has been taking shape over the past 30 years, the origins of which go back at least to the 1930s. As the name suggests, and as described by Joseph Mahoney, much of the Economics of Organization is interdisciplinary—drawing, as it does, on contributions from economics, organization theory (broadly conceived to include sociology, political science, and social psychology), cognitive psychology, and aspects of the law, especially property and contract law.

In large measure, the Economics of Organization arose in response to perceived limitations in the neoclassical theory of the firm, which described a firm as a production function that transformed inputs into outputs according to the laws of technology. Useful as this construction was (and is) for the study of prices and output (and, more generally, of resource allocation) this theory was less useful for understanding the modern corporation, and for the making of public policy. The seeds of discontent were planted by Adolph Berle and Gardiner Means (1932) in relation to the separation of ownership from control, by Ronald Coase (1937) with reference to the puzzle of vertical integration and the theory of the firm, by Chester Barnard (1938), who featured adaptation of a cooperative kind accomplished through hierarchy (as opposed to adaptation through the market in response to changes in relative prices), by the ‘marginalist controversy’ of the 1940s, and by Armen Alchian’s (1950) appeal to evolutionary arguments to explain observed economic regularities.

The short of it is that too much was being asked of the neoclassical theory of the firm. As Harold Demsetz put it, it is ‘a mistake to confuse the firm of [orthodox] economic theory with its

real world namesake. The chief mission of neoclassical economics is to understand how the price system coordinates the use of resources, not the inner workings of real firms' (1988, p. 189). Efforts to use the neoclassical theory of the firm in an all-purpose way nevertheless persisted. Strange and convoluted interpretations of nonstandard and unfamiliar contracting practices and organizational structures often resulted, which interpretations were sometimes responsible for public policy error.

A new theory of the firm was evidently needed. But as Mahoney makes clear, what we have is not a new theory (singular) but new theories (plural). The five new theories of the firm (more generally, of economic organization, in that the firm is not a stand-alone entity but is to be examined in relation to the market and other modes of organizing economic activity) described by Mahoney are: the behavioral theory of the firm, transaction costs theory, property rights theory, agency theory, and resource-based/dynamic capabilities. In many respects these are complementary, in that they illuminate different issues. But sometimes they collide, in that they provide rival explanations for the same phenomena. Ways have nevertheless been found to join parts of these, and more of this is in prospect. As matters stand presently, the study of economic organization is at a pre-unified state of development.

Readers of this book will want to ask themselves, as they work their way through successive chapters, what are the strengths and weaknesses of each proposed perspective? What phenomena does each illuminate? What predictions? What is the implied empirical research agenda? Does the theory scale up from the simple (often two actor) model to complex (many actor) economic organization? What public policy ramifications accrue?

The new theories of the economics of organization that have taken shape over the past thirty years have vastly expanded the research agenda in both economic theory and organization

theory as well as in the applied fields to which Mahoney makes reference: strategic management, agricultural economics, industrial organization, international business, and marketing. Indeed, applications to business and economic history, economic development, transition economics, positive political theory, and the law have also been made and are in progress. The economics of organization is an idea whose time has come. Lucky are the students of economic organization who seize upon the opportunity to run with this proliferation of good ideas.

Oliver E. Williamson

January 2003



## **Preface**

I have many people to thank on my journey to writing this book. I begin with my mother who advised me to “be kind.” My father advised me that: “No matter what happens to you in life, no one can take your education from you.” My younger sister Rose taught me courage and my younger brother Jim, who in many ways I looked up to during our childhood days, taught me caring and thoughtful reflection.

In terms of education, I first thank all of my grade school and high school teachers who showed dedication to their chosen profession. Studying undergraduate economics in the Faculty of Arts and Sciences at the University of Pennsylvania provided a wonderful foundation that has influenced my thinking and I especially thank Professors Jacques Cremer, Robert Inman, Irving Kravis, Laurence Seidman, and Sidney Weintraub. In particular, Sidney Weintraub’s passion for ideas has had a lasting influence on my life.

In graduate school studying Business Economics at the Wharton School of Business of the University of Pennsylvania, I thank Professors Robert Pollak and Harbir Singh for teaching me microeconomic theory, and corporate strategy, respectively. I thank Professors Claudia Goldin, Bruce Kogut, and Gordon Walker for serving on my doctoral committee. My dissertation advisor, Professor Almarin Phillips provided encouragement and guidance. Professor Ned Bowman not only served on my dissertation committee, but also was a source of inspiration for me both personally and professionally. Support from the Reginald Jones Center, under Ned’s guidance, is gratefully acknowledged. I thank Professor Richard Marston, who gave me wise counsel to work for Ned Bowman at Wharton.

At the University of Illinois at Urbana-Champaign I thank my numerous colleagues over the past 15 years. I especially thank Irene Duhaime, Anne Huff, and Ravi Madhavan for their friendship and support. I also thank the excellent doctoral students that I have worked with over the years such as Bill Bogner, Jongwook Kim, Yasemin Kor, Sung Min Kim, Chamu Sundaramurthy, and Danchi Tan, among others.

Finally, I owe so much to my wife, Jeanne Marie Connell, who has been there in good times and in bad. She has been a guiding force and a source of constant love and compassion. To her I dedicate this book. Thank you, Jeanne.

## **Overview**

While writing this research book on the economic foundations of strategy, I kept two issues in mind. Suppose readers (at various levels of previous training) wanted to learn about organizational economics, especially from the perspective of strategic management: What significant themes best capture the directional tendencies of organizational economics today? How did current positions evolve? Suppose readers wanted to develop their own capacity to theorize and carry out research within the tradition of the economics of organization: What unresolved issues in this field are especially relevant to modern strategy research? What concepts are key building blocks? Where is additional empirical evidence needed?

Organizational economics has been informed by many great minds in social science research including Nobel Prize recipients in economics such as Kenneth Arrow, Ronald Coase, Douglass North, and Herbert Simon, and potential future recipients such as Oliver Hart (for property rights theory), and Oliver Williamson (for transaction costs theory). In addition, research contributions by business school professors such as Alfred Chandler, James March, and Sidney Winter provide a wealth of insights based upon business experiences. It is an important field taking a central place in the study of strategy, though of course it has also been useful for policy and other purposes (such as agricultural economics, international business studies, management information systems, marketing, organization theory, and so on).

In this book, *organizational economics* includes the following five inter-related theories: (1) a behavioral theory of the firm; (2) transaction costs theory, (3) property rights theory; (4) agency theory; and (5) (evolutionary) resource-based theory. In each area I have summarized my view of the critical observations of a few authors who have shaped the theory. Their work is

presented in the present tense --- even though a number of contributors unfortunately are no longer alive --- because the work itself is very much alive. These books are worthy of careful attention. Not only are these books widely cited, they exemplify the way theories are constructed by individual authors, and the way the conversation among authors develops over time to create complex and compelling answers to important questions.

In the conclusion, I discuss complementarities among the theories outlined in this book, while recognizing their distinctive features. Most importantly, I outline some important questions for the ongoing economic study of organization that draws on these resources. I have written this book because the seminal works cited here are the foundations for so much current research. Key contributions of each author, in my view, are emphasized in italics. It may surprise some readers to find how current these observations sound, and some links to other research literatures are discussed here. But that discussion is limited to accommodate the wide variety of directions these foundational works can support. I hope that many different readers will be as inspired as I am by this classic literature. By understanding and emulating the best of our predecessors we can make more significant contributions today.

Joseph T. Mahoney

November 2003

## **Chapter 1 Behavioral Theory of the Firm**

The chapter begins with Barnard's (1938) *The Functions of the Executive*, and is followed by four books from the 'Carnegie School:' Simon's (1947) *Administrative Behavior*; March and Simon's (1958) *Organizations*; Cyert and March's (1963) *A Behavioral Theory of the Firm*; and Simon's (1982) *Models of Bounded Rationality: Behavioral Economics and Business Organization*. These books contain some of the best scholarly writings that the research literature has to offer on the behavioral theory of the firm. These research books are worth studying in detail because they continue to be widely cited today, and because their clarity and relevance have not yet been surpassed.

The decision to classify the behavioral theory of the firm as part of an organizational economics approach to strategic management has its precedents, notably in the work of Barney and Ouchi (1986). Nonetheless, given that my book is part of a series, it might seem that this topic should be reserved for a research book on the Behavioral Foundations of Strategy. I am sure it will reappear in that context. However, the behavioral theory of the firm also is part of organizational economics. Organizational economics is a multi-disciplinary endeavor that draws on the broader field of economics but also gives attention to contributions from organization theory, law, and other areas. As an important example, Herbert Simon, whose 1947 and 1982 books are reviewed in this chapter, was awarded the Nobel Prize in the discipline of economics for work that included his contributions to the behavioral theory of the firm.

Furthermore, the behavioral theory of the firm serves as an important building block in transaction costs theory (Williamson, 1975). This theory is the subject of Chapter 2, and a central topic in organizational economics. Behavioral theory is also an important building block in

dynamic capabilities theory and evolutionary economics (Nelson and Winter, 1982). This research is the subject of chapter 5.

In terms of the five books chosen, Barnard (1938) combines the two cultures of science and art and it is the aesthetic reading of Barnard (1938) that explains the intensity of students' responses to this work. Barnard (1938) offers an intense, structured and coherent art form that depends on students using their capacities and readiness to apprehend the aesthetic experience of management based on the author's intimate, habitual, interested experience (Mahoney, 2002).

Simon (1947) proposes a theory of human choice and decision-making that aims to accommodate both those rational aspects of choice that have been the principal concern of economists and those properties and limitations of the human decision-making mechanisms that have attracted the attention of psychologists and practical decision-makers. Simon (1947) focuses primarily on the decision-making processes that are internal to the organization. Simon (1947) describes how organizations influence the decisions of their members, bring about consistency among those decisions, and guarantee that the decisions will be compatible with the overall organizational goals.

March and Simon (1958) persuasively argue that an adequate study of human behavior in organizations must take into account the motivational, attitudinal, and rational aspects of human behavior. Thus, both the works of economists on the planning process and the works of psychologists on organizational communication and problem-solving capabilities contribute to the evolving science of organization.

Cyert and March (1963) emphasize the actual process of making business decisions and provide detailed observations of the ways in which organizations make these decisions. Cyert and March (1963) develop an empirically relevant, process-oriented general theory of economic

decision making by a business firm that, in my judgment, has stood the test of time. Cyert and March (1963) present the rudiments of a behavioral theory of the firm that have proven to be relevant both to economic theory and to the theory of complex organizations.

Simon (1982) take up where *Administrative Behavior* (1947) left off --- attempting to understand decision making in its most general sense and, in particular, to show that economics and psychology could contribute to illuminating organizational decision-making processes. More specifically, Simon (1982) is concerned with explaining why there has been so little mutual influence of economics and psychology upon each other, why a deeper dialogue needs to be developed between these two disciplines, and what the subject matter of their discourse could be.

In the process, Simon (1982) reveals a deep belief in and commitment to the interdependencies and complementarity of the several social sciences. Simon (1982) borrows not only from economics, but also from operations research, artificial intelligence, and cognitive psychology for the purpose of building a theory of procedural rationality (i.e., a theory of the processes of decision-making) in complex, dynamic circumstances.

Though these arguments are a sufficient introduction to the chapter, I would make a final observation related to teaching. Those in strategic management who teach managers and managers-to-be will know that our students appreciate receiving not only theories for predicting but also theories that provide explanation. In other words, practitioners appreciate know-how, but are deeply seeking advances in know-why. It has been my experience in teaching executives that a *Behavioral Theory of the Firm* resonates with these managers and proves instructive for them. It makes beginning with the work of a practicing manager highly appropriate.

Barnard, Chester I. (1938). The Functions of the Executive. Cambridge, MA: Harvard University Press.

In my judgment, this book is the most high-powered intellectual contribution to organization or economic theory ever written by a practicing manager. Barnard's (1938) purpose is to provide a comprehensive theory of cooperative behavior in formal organizations.<sup>1</sup> Barnard (1938) observes that formal organization involves conscious, deliberate, and purposeful cooperation among people. One of the indispensable functions of an organization is to promote communication among these individuals. Another function is to maintain cohesiveness by regulating the willingness of various stakeholders to serve the organization, and by maintaining the stability of authority. A third function is to maintain a feeling of personal integrity, of self-respect, and of independent choice.

But Barnard (1938) maintains that successful cooperation in or by formal organizations is the abnormal, not the normal condition. We observe from day to day the successful survivors among innumerable organizational failures. Failure to cooperate, failure of cooperation, failure of organization, disorganization, dis-integration, destruction of organization – and reorganization – are the characteristic facts of human history.

The executive is critical. Executives inculcate belief in a common purpose. More concretely, executives synthesize the actions of contradictory forces and reconcile conflicting instincts, interests, conditions, positions, and ideals.

Informal Organization. While Barnard (1938) defines the formal organization as a system of consciously coordinated activities or forces of two or more persons, this book also

---

<sup>1</sup> In addition to Barnard's (1938) classic, Barnard (1948) provides a collection of his selected papers. For modern assessments of Barnard (1938), see Mahoney, Huff and Huff (1994a, 1994b), Mahoney (2002), Scott (1987), and Williamson (1995). Mahoney (2002) summarizes the major elements of Barnard's

emphasizes the important role of informal organization within formal organizations. Crucially, Barnard (1938) regards informal organization as a means of maintaining the personality of the individual against certain effects of formal organizations that tend to disintegrate the personality. In fact, Barnard concludes that *expansion of cooperation and the development of the individual are mutually dependent realities, and that a due proportion or balance between them is a necessary condition of human welfare.*

Incentives. Barnard (1938) observes that incentives are fundamental in formal organization. Inadequate incentives mean dissolution, unwarranted changes of organization purpose, or failure of cooperation. Hence, in all sorts of organizations, affording adequate incentives becomes essential. The specific means available include: (a) material inducements, not just money but other things, (b) personal, non-material inducements including distinction, prestige and personal power, (c) desirable physical conditions, (d) 'ideal benefactions' by which Barnard (1938) means the capacity of organizations to satisfy personal ideals.

The remarks about personal ideals and interests are very much in line with more recent discussions about identity and identification. Barnard (1938) is also contemporary in recognizing the incentives associated with (e) social attractiveness, or the social compatibilities people feel in their work environment, (f) conditions of habitual methods and attitudes, (g) the opportunity for enlarged participation and (h) the condition of communion, or feeling of solidarity or comradeship. None of this solidarity happens without effort. In addition to incentives, the book discusses persuasion, and the inculcation of motives as important aspects of the organization.

---

(1938, 1948) theory with special attention to Barnard's concepts of leadership and responsibility that are



Authority. Authority is the character of a communication (or order) in a formal organization by virtue of which a contributor accepts such an order. Barnard (1938) suggests that a person can and will accept a communication as authoritative only when four conditions simultaneously obtain:

- The person can and does understand the communication;
- At the time of the person's decision the person believes that the order is not inconsistent with the purpose of the organization;
- At the time of the person's decision, the person believes the order to be compatible with his or her personal interest as a whole; and
- The person is able mentally and physically to comply with the order.

Perhaps the most well known idea in the book is found in this discussion. Barnard (1938) argues that *there exists a 'zone of indifference' in each individual within which orders are acceptable without conscious questioning of their authority.* Barnard (1938) further maintains that since the efficiency of organization is affected by the degree to which individuals assent to others, denying the authority of an organization communication is a threat to the interests of all individuals who derive a net advantage from their connection with the organization, unless the orders are unacceptable to them also. Thus, nothing is more 'real' than 'authority.'

An interesting corollary can be found in the assertion that the fine art of executive decision-making includes not deciding questions that are not now pertinent, not deciding prematurely, not making a decision that cannot be made effective, and not making decisions that others should make. These are interesting, and rather unique observations; Barnard (1938) argues the proper use of authority preserves morale, develops competence, and maintains authority.

---

essential for distinctive competence (Selznick, 1957).

However, the natural reluctance of some people to decide, their persistent disposition to avoid responsibility, and their fear of criticism, typically overwhelm executives. Executives thus must learn to protect themselves from the excessive burdens of decision-making, if they are not already protected by a well-regulated and habitual distribution of responsibilities.

Another contemporary feature of this book is that the executive process Barnard (1938) describes transcends intellectual methods. ‘Feeling,’ ‘judgment,’ ‘sense,’ ‘proportion,’ ‘balance,’ ‘appropriateness’ and other words are used to describe what executives should aspire to becoming. Leadership is more a matter of art than a matter of science. The processes used are more aesthetic than logical, derived chiefly from ‘intimate,’ ‘habitual,’ ‘interested’ experience. For Barnard (1938), *coordination is a creative act*.

Executive responsibility is also emphasized. Whatever morality exists in an individual becomes effective in his or her conduct, and the organization as a collective of cooperating individuals endures in proportion to the breadth of the morality by which it is governed. This assertion is only to say that foresight, long purposes, and high ideals are the basis for the persistence of cooperation (e.g., “old men and old women plant trees”).

Although emphasizing instincts and morality, Barnard (1938) believes that a ‘science of organization’ is also possible. Barnard (1938) recommends that treatises on management be written from various perspectives, including social anthropology, sociology, social psychology, and institutional economics. But Barnard (1938) warns that we should not deceive ourselves by thinking that a science of cooperation and organization will alone promote greater integration of social forces. Inspiration is necessary to inculcate the sense of unity, and to create economic ideals. Emotional rather than intellectual acceptance is required.

Barnard (1938) presents a *systems* view of the organization that contains a psychological theory of motivation and behavior, a sociological theory of cooperation and complex interdependencies, and an ideology based on a meritocracy. These insights greatly influenced Simon (1947), to whose early and influential book we now turn.

Indeed, Barnard wrote the foreword to Simon's (1947) *Administrative Behavior*. Barnard writes that: "[Simon's book] has the right 'feel.' This means that I find Professor Simon's apprehension of the structure of organized action consonant with my own experience. It therefore appeals to me as sound" (1947: xliii). From Simon's (1947) classic book concerning decision-making processes, readers should be able to discern principles of general organization that apply to administrative organization of great variety. Simon (1947) provides us with a self-conscious attempt to develop adequate linguistic and conceptual tools for realistically and significantly describing organizations. Simon's (1947) primary thesis is that decision-making is the heart of organization, and that the vocabulary of organization theory must be derived from the logic and psychology of human choice.

Simon, Herbert A. (1947). Administrative Behavior. New York: Free Press.

Simon (1947) provides a brilliant synthesis of the practical teachings of Barnard (1938) and the evolving positive science of organization theory. As already noted, Simon (1947) is a landmark in organization theory as well as the economics of organization. Indeed, the organization theorist William Scott (1987: 45) classifies Simon (1947) within the paradigm of ‘organizations as rational systems.’ From the perspective of the rational systems view, the behavior of organizations is considered as actions performed by purposeful and coordinated agents. In this sense, Simon (1947) is consistent with the logic of economics, and uses the familiar language of *information, efficiency, implementation, and design*. Unlike neoclassical economics, however, Simon (1947) also insists on coming to terms with cognitive limitations, which are discussed in terms of *constraints, authority, routines, and bounded rationality*. These terms imply that the rationality of organization behavior takes place within clearly specified limits. In short, this landmark book provides *an attention-based theory of the firm* of interest to both economic and organizational theorists.

Bounded Rationality. Simon (1947) observes that a person does not live for years in a particular position in an organization, exposed to some streams of communication, shielded from other streams of communication, without profound effects upon what the person knows, believes, hopes, emphasizes, fears, and proposes. Researchers can understand neither the ‘input’ nor the ‘output’ of executives without understanding the organization in which executives work.

The term *organization*, for Simon (1947), refers to a complex pattern of human communications and relationships. This pattern of relationships provides each member of an organization or group within an organization much of the information and many of the

assumptions, goals, and attitudes that enter into decisions. The pattern of relationships provides a set of stable and comprehensible expectations as to what the other members of the group are doing and how other members are likely react to what is said and done. Every executive makes decisions and takes actions with one eye on the matter itself and one eye on the effects of *this* decision upon the future pattern of relationships -- that is to say, upon its organizational consequences.

In summary, organizations are important because they provide much of the input that develops an executive's personal qualities and habits. Organizations also provide those in responsible positions the means for exercising authority and influence over others, a topic discussed in some detail in the following pages. Third, the organization influences the environments of information in which decisions are carried out.

When executives give attention to these indirect consequences, they concern themselves with organization. Sales managers react like sales managers because they occupy particular organizational positions, receive particular kinds of communications, are responsible for particular sub-goals, and experience particular kinds of (economic) pressures. Executives can modify beliefs and attitudes by changing the flows of communications, and thus modify decisions being made.

Decisions are also influenced by the authority relationship. On the one hand, classical organization theory emphasizes formal lines of authority in a hierarchical organization, implying (as Barnard observed) that legitimate commands are typically carried out. On the other hand, the 'human relations' school emphasizes the value of broad participation in decision-making, demonstrates the importance of informal organization and the consequent limits on formal

authority, and raises difficult questions about the human costs of excessively authoritarian environments.

These are not only different, more social, concepts of authority, but also different concepts of rationality. Simon (1947) argues that the social sciences suffer from acute schizophrenia in their treatment of rationality. At one extreme, neoclassical economists attribute to 'economic man' an omniscient rationality. Economic man has a complete and consistent system of preferences that allows him to choose correctly among the entire set of alternatives available to him. He is completely aware of what these alternatives are, there are no limits on the complexity of the computations he can perform in order to determine what alternatives are best, and he correctly makes all probability calculations. Tendencies uncovered by social psychology, traceable to Freud, which tend to reduce all cognition to affect, are at the other extreme. This alternative point of view notes that coins look larger to poor children than to rich children, observes that the pressures of a social group can persuade a person that he or she sees spots that are not there, shows that the process of group problem solving involves accumulating and discharging tensions, and so on.

Simon's (1947) major contribution to the economics of organization, as well as to organization theory, is the argument that it is precisely in the realm where human behavior is intendedly rational, but only limitedly so, that there is room for a genuine theory of organization. Simon (1947) maintains that *organizational behavior is the theory of intended and bounded rationality – it is about the behavior of humans who satisfice because they do not have the abilities to maximize. Whereas neoclassical 'economic man' maximizes -- selects the best alternative from among all those available to him, 'organizational man' satisfices -- looks for a*

*course of action that is satisfactory or 'good enough.' Economic man deals with the 'real world' in all of its complexity. The world organization man perceives is a drastically simplified model of the real world.*

What is the significance of these two characteristics of satisficing and bounded rationality for organizational man? First, because he satisfices rather than maximizes, organizational man can make choices without first examining all possible behavior alternatives, and without ascertaining that alternatives considered are in fact all those available. Second, because he ignores the interrelatedness of all things (so stupefying to thought and action), organizational man can make decisions with relatively simple rules of thumb that do not make impossible demands upon the capacity for thought. These critical theoretic observations have many interesting consequences.

Decision-Making and Administrative Organization. Simon (1947) argues that it is *the process of choice that leads to action*. Although any practical activity involves both 'deciding' and 'doing,' it was not commonly recognized until this important book that a theory of organization should be concerned with the processes of decision as well as with the processes of action.

Simon (1947) notes that all behavior involves conscious or unconscious selection of particular actions out of all those actions that are physically possible to the actor and to those persons over whom the actor exercises influence and authority. The term 'selection' is used without any implication of a conscious or deliberate process. Selection refers to the fact that, if the individual follows one particular course of action, there are other courses of action that the individual thereby forgoes.

Simon (1947) employs a definition of authority substantially equivalent to that put forth

by Barnard (1938). Subordinates accept authority whenever these subordinates permit their behaviors to be guided by the decision of a superior, without independently examining the merits of that decision. When exercising authority, the superior does not seek to convince the subordinates, but only to obtain their acquiescence. In actual managerial practice, of course, authority is usually mixed with suggestion and persuasion. If a superior attempts to carry authority beyond a certain point, which may be described as the subordinate's 'zone of acceptance,' disobedience will typically follow.

The magnitude of the zone of acceptance depends upon the various sanctions which authority has available to enforce its commands. The structure of formal authority in an organization typically is related to the appointment, disciplining, and dismissal of personnel. Informal authority relations in the tactical (day-to-day) work of the organization commonly supplement these formal lines of authority. The formal hierarchy is largely reserved for hearing and settling disputes.

Problems of Organizational Theory. Simon (1947) maintains that the authority relationship enables an organization to bring about specialization in the work of making decisions, so that each decision is made at the point in the organization where the decision can be made more expertly for achieving purpose. 'Purpose' is defined as the objective for which an activity is carried on, and 'process,' as a means of accomplishing purpose.

Simplistically, the concept of *purpose* involves a hierarchy of decisions -- each step downward in the hierarchy consisting of an implementation of the goals set forth in the step immediately above. Behavior is purposive in so far as it is guided by general goals; it is rational in so far as it selects alternatives that are conducive to the achievement of the previously selected



goals. More realistically, the achievement of purpose often requires attention along multiple dimensions in the organization. Providing a useful analogy to make this particular point, Simon (1947) states that closet space is an important item in the design of a successful house, yet a house designed entirely with a view to securing a maximum of closet space -- all other considerations being forgotten -- would be considered somewhat unbalanced.

Similarly, unity of command, specialization by function, and decentralization are items to be considered in the design of an efficient organization. No single item is of sufficient importance to suffice as a guiding principle for the organizational analyst. In the design of organizations, as in their operation, overall efficiency is a guiding criterion. Mutually incompatible advantages are balanced against each other, just as an architect weighs the advantages of additional closet space against the advantages of a larger living room. A valid approach to the science of organization requires that the relevant diagnostic criteria be identified; that each organizational situation be analyzed in terms of the relevant set of criteria; and that research be instituted to determine how weights can be assigned to the several criteria when they are mutually incompatible.

Simon (1947) also observes that before a science of organization can develop theoretical principles, it must possess concepts. To be scientifically useful, the concepts chosen must be operational; that is, their conceptual meanings must correspond to empirically observable facts. As an analogy, before a law of gravitation could be formulated, it was necessary to have the concepts of 'acceleration' and 'weight,' and there had to be commonly accepted measures of these terms.

The theory of organization, in Simon's (1947) view, is concerned with how an enterprise should be constructed and operated in order to accomplish its work efficiently. A fundamental

principle of organization, which follows almost immediately from the rational character of 'good' organization, is that among several alternatives involving the same expenditure the one should be selected that leads to the greatest accomplishment of organizational objectives; and among several alternatives that lead to the same accomplishment the one should be selected that involves the least expenditure. Since this 'principle of efficiency' is characteristic of any activity that attempts rationally to maximize the attainment of certain ends with the use of scarce means, it is as characteristic of economic theory as it is of organizational theory. In this sense, the 'organization man' takes his place alongside the neoclassical 'economic man.' However, as already noted, one of Simon's (1947) most important contributions to science is the argument that individuals are limited by those skills, habits, and reflexes that are no longer in the realm of the conscious.

Limits to Rationality. At the most simplistic level, performance may be limited by manual dexterity or reaction time, and decision-making processes may be limited by the speed of mental processes. Individuals also are limited by their values and those conceptions of purpose that influence them in making their decisions, and these tend to be shaped by their organizational experience. If their loyalty to the organization is high, their decisions may evidence sincere acceptance of the objectives set for the organization; if loyalty is lacking, personal motives may interfere with organizational efficiency. Finally, individuals are limited by their knowledge of factors relevant to their job. This limitation applies both to basic knowledge required in decision-making -- bridge designers must know the fundamentals of mechanics -- and to the information that is required to make decisions appropriate to a given situation.

In discussing means and ends, as well as facts and values, Simon (1947) is forthcoming concerning his own (logical positivist) philosophical perspective. Simon (1947) maintains that every decision involves elements of two kinds, which are called 'factual' and 'value' elements respectively. This distinction is of primary importance for organization. Simon (1947) holds as a fundamental premise the idea that ethical terms are not completely reducible to factual terms. There is therefore no way in which the correctness of ethical propositions can be empirically tested. From this positivist perspective, if a sentence declares that some particular state of affairs 'ought to be,' or that it is 'preferable' or 'desirable,' then the sentence performs an imperative function, and is neither true nor false.

In contrast, a statement concerning the observable world is factual if, in principle, its truth or falsity may be tested. Simon (1947) hastens to add that in practice, the separation between ethical and factual elements in judgment can usually be carried only over a short distance. Further, the values involved in organizational decisions are seldom the final values in any psychological or philosophical sense.

Rationality in Organizational Behavior. Rationality in the world of experience is a complex concept. Simon (1947) provides a scenario of two soldiers who sit in a trench opposite a machine-gun. One soldier stays under cover. The other soldier, at the cost of his life, destroys the machine-gun with a grenade. Which action is rational? *Simon (1947) suggests that perhaps the only way to clarify these complexities is to use the term 'rational' in conjunction with appropriate adverbs.* Action may be called 'objectively' rational, if *in fact*, it is the correct behavior for maximizing given values in a given situation. An action is 'subjectively' rational if it maximizes attainment relative to the actual knowledge of the subject. An action is 'consciously' rational to the degree that the adjustment of means to ends is a conscious process.

An action is 'deliberately' rational to the degree that the adjustment of means to ends has been deliberately brought about (by the individual or by the organization). An action is 'organizationally' rational if it is oriented to the organization's goals; an action is 'personally' rational if it is oriented to the individual's goals.

The Psychology of Organizational Decisions. From a rational point of view choice is the process by which an alternative for each moment's behavior is selected. The task of choice involves three steps: (1) the listing of all alternatives; (2) the determination of all the consequences that follow upon each of these alternatives; and (3) the comparative evaluation of these sets of consequences. Each individual, in order to determine uniquely the consequences of actions, must know what will be the actions of others. This knowledge is of fundamental importance for the whole process of decision-making.

Things are not so simple from an organizational point of view. Simon (1947) agrees with Barnard (1938) that organizations are systems of cooperative behavior. From the logical positivist perspective, rationality concerns the selection of preferred behavior alternatives in terms of some system of values whereby the consequences of behavior can be evaluated. But Simon (1947) argues that it is impossible for the behavior of a single, isolated individual to reach a high degree of rationality. The number of alternatives the individual must explore is so great, the information the individual would need to evaluate so vast that even an approximation to objective rationality is hard to conceive. Individual choice takes place in an environment of 'givens' -- premises that are accepted by the individual as bases for choice; and behavior is adaptive only within the limits set by these 'givens.'

Objective rationality would imply that the behaving individual molds behavior into an integrated pattern by (a) viewing the behavior alternatives prior to choice in panoramic fashion, (b) considering the whole complex of consequences that would follow from each choice, and (c) with the system of values as criterion singling out one from the whole set of alternatives. Observed behavior, even that which is ordinarily considered 'rational,' possesses many elements of disconnectedness not present in this idealized picture. However, one function the organization performs is to place members in a psychological environment that helps adapt their choices to organizational objectives. It also provides them with information needed to make these choices.

Even so, *if individual or organization behavior is examined over time, it exhibits a mosaic character. Each piece of the pattern may tend to be integrated with others by their orientation to a common purpose; but such purpose shifts from time to time with shifts in knowledge and attention, and is held together in only slight measure by any conception of an overall criterion of choice.* Actual behavior falls short, in at least two ways, from objective rationality. First, rationality requires both complete knowledge and total anticipation of the consequences that will follow each choice. In fact, knowledge of consequences is always fragmentary. Second, rationality requires a choice among all possible alternative behaviors. In actual behavior, only a few of all these possible alternatives come to mind. Complete rationality is limited by lack of knowledge.

Humans striving for rationality but restricted within the limits of their knowledge develop working procedures that partially overcome these difficulties. These procedures assume they can isolate from the rest of the world a closed system containing only a limited number of variables and a limited range of consequences. Simon (1947) notes that the problem of discovering what factors are, and what are factors are not, important in any given situation is as essential to choice

as knowledge of the empirical laws governing those factors that are finally selected as relevant. *Rational choice is feasible to the extent that the limited set of factors upon which decision is based corresponds, in nature, to a closed system of variables, that is, to the extent that significant indirect effects are absent.* Only in the cases of some important decisions is it possible to bring to bear sufficient resources to unravel an involved chain of effects.

Humans do have some important abilities, including the capacity to observe regularities in nature, and to communicate with others. Both help to shorten materially the learning process. The first capability means that previous experiences with other choices (of the same sort) may enable decision-makers to infer something about the character of the particular choice that they face. Then, communication provides a tremendous advantage in learning. For example, engineers designing a pavement do not have to base their attempts entirely upon experimentation, but can use reference sources that describe the conclusions of others. In effect, a relatively small amount of experience can serve as the basis for a wide range of decisions.

*Memory.* Memory as described in this book may be either natural or artificial -- information may be stored in the mind, or it may be recorded in such a way as to be accessible. The artificial kind of memory that consists of libraries, files, and records, is the most important in organizations. For either natural or artificial memory to be useful, there must be mechanisms that enable the memory to be drawn upon when needed. Hence, human rationality relies heavily upon the psychological and artificial associational and indexing devices that make the store of memory accessible when needed for the making of decisions.

An equally important mechanism that assists in the preservation of useful behavior patterns is habit. Habit, like memory, has an artificial organization counterpart that can be termed

‘organizational routine.’ This idea has become important in the more recent literature, notably Nelson and Winter’s (1982) work reviewed in Chapter 5.

*Attention.* Memory affects attention, where attention refers to the set of elements that enter into consciousness at any given time. To a considerable extent, the limits of rationality are the result of the limits of attention. It is important that both attention and behavior, once initiated in a particular direction, tend to persist in that direction for a considerable time interval. This persistence of attention holds even when the original choice of activity was a matter of relative indifference. Activity often results in psychological ‘sunk costs’ that make persistence of attention in the same direction advantageous. A second reason for persistence is that the activity itself creates stimuli that focus attention toward its continuance and completion.

Simon (1947) maintains that the process of the integration of behavior involves three principal steps:

- Individuals (or organizations) makes broad decisions regarding the values to which they are going to direct their activities for *substantive planning*;
- Individuals design and establish mechanisms that will direct their attention, and channel information and knowledge in such a way as to cause the specific day-to-day decisions to conform to the substantive plan. This decisional activity is called *procedural planning*; and
- Individuals execute the plan.

But there are at least two intervening organizational and institutional influences on individual behavior: First, organizations and institutions permit (indeed encourage) stable expectations, and second, organizations and institutions provide the general stimuli and attention-directors that channel the behaviors of members of the group, and provide those members with the intermediate goals that stimulate action.

Simon (1947) suggests the following mechanisms of organization influence:

- (1) The organization divides work among its members. By giving each worker a particular task to accomplish, it directs and limits attention to that task;
- (2) The organization establishes standard operating procedures;
- (3) The organization transmits decisions by establishing systems of authority and influence;
- (4) The organization provides (formal and informal) channels of communication running in all directions through which information flows; and
- (5) The organization trains and inculcates its organizational members.

The Equilibrium of the Organization. Simon (1947) maintains that individuals are willing to accept organization membership when their activity in the organization contributes, directly or indirectly, to their personal goals. The phrase 'personal goals' should be understood in a broad sense. It is not restricted to egoistic goals, much less to economic goals. In a discussion reminiscent of Barnard (1938), the members of an organization contribute to the organization in return for *inducements* that the organization offers them. If the sum of the *contributions* is sufficient to supply the necessary inducements, the organization survives and grows; otherwise, the organization shrinks and ultimately disappears. In return for their inducements, members typically offer the organization not a specific service but their undifferentiated time and effort. Organizational members place this time and effort at the disposal of those directing the organization, to be used as those directing see fit. Thus, both the customer relation (in the commercial organization) and the employee relation originate in contract, but in contracts of different kinds. The employment contract results in the creation of an authority relationship between the organization and the employee.



How can this be? Why does the employee sign a blank check, so to speak, in entering upon employment? First, from the perspective of the organization, nothing would be gained by offering inducements to employees unless the employees' behaviors could be brought into a system of organization behavior through their acceptance of its authority. Second, from the perspective of the employees, the precise activities with which their time of employment is occupied may, within certain limits, be a matter of relative indifference to them. In addition to the salary that employees receive, employees may value the status and prestige that their positions in the organization give them, and employees may value their relations with the working group of which they are part.

Organizational Goals. Three bricklayers were asked what they were doing. 'Laying bricks,' 'Building a wall,' 'Helping to build a great cathedral,' were their respective answers. This story conveys Simon's (1947) idea that in the world of experience the line of demarcation between personal and professional interests is not a sharp one, for personal satisfactions may arise from the competent performance of a professional role, and both personal satisfactions and dissatisfactions may arise from innumerable conditions that surround the employment relationship.

Particular professional training may provide individuals with specific techniques and knowledge for solving problems (e.g., accounting techniques, legal techniques, and so on), which are then drawn upon as part of the program evoked by their roles. In this way, a chief executive with an accounting background may find different problem solutions from a chief executive, in the same position, with a legal background. Individuals may incorporate in their role not only a professional style but also a personal style. Individuals may bring to the role, for example, habits and beliefs that provide them with crucial premises for their handling of interpersonal relationships.

An interesting question is why most commercial organizations tend to maintain fairly stable goals. Simon's (1947) answer is, first, that there are both economic and psychological 'sunk costs' that make rapid adjustment unprofitable. Second, the organization requires know-how in a particular field – which becomes a 'sunk asset' and part of the influencing organizational environment. Third, the organization acquires goodwill, which is also a 'sunk asset' (i.e., the asset is not easily re-deployed) and, thus, is not readily transferable to another area of activity.

The Role of Authority. 'Authority' is defined as the power to make decisions that guide the actions of another. It is a relationship between two individuals, one 'superior' and the other 'subordinate.' The superior transmits decisions with the expectation that the subordinate will accept these decisions. The subordinate expects such decisions, and these decisions influence the conduct of the subordinate. The relationship of authority is defined, therefore, in behavioral terms. Authority involves behaviors on the part of both superior and subordinate. When, and only when, these behaviors occur does an authority relationship exist between the superior and subordinate.

Individuals who do not have recognized status, or who are not recognized by their associates as expert with respect to a certain kind of knowledge, will have a more difficult time convincing their listeners that a recommendation is sound than those who possess the credentials of 'expertness.' Recommendations are judged partly on their merits, but are judged partly on the expertise of the persons making the recommendations. This pattern of judgments holds both because the individuals acting upon the recommendations often do not have the expertise needed

to judge them, and because time pressures require these individuals to accept the recommendations of those whom they trust.

Furthermore, it is not implied that this resistance to 'irregular' suggestions is entirely a weakness of organization. The specialization of decision-making functions, and the fixing of responsibility for particular kinds of expertness upon particular individuals, is an important source of organizational efficiency that needs to be balanced against the potential loss of independent ideas that results. When there is a disagreement between two persons, and when the disagreement is not resolved by discussion, persuasion, or other means of conviction, then the disagreement must be decided by the authority of one or the other participant. It is this 'right to the last word' that is usually meant in speaking of 'lines of authority' in an organization.

Simon (1947) proposes that the degree of obedience expected will vary with the social situation. The American workers of his day, for example, probably had a somewhat wider zone of acceptance, so far as the employer's instructions are concerned, than workers today. In part this difference in the degree of authority may be due to the worker's weaker bargaining position then, or conversely, the stronger sanctions of the employer; but there is probably a more fundamental change in social attitudes as to what is 'proper' for an employer to ask an employee to do. These changed attitudes are reflected in social legislation limiting the terms of the employment contract. Professional workers and skilled workers are apt to have relatively narrow zones of acceptance, particularly in the areas of their own professional competencies.

The field of organizational behavior has stressed 'purpose' as a sanction of primary importance. Subordinates are willing to obey commands because subordinates realize that the coordination secured thereby is useful to the attainment of the joint purpose. Several conditions must be satisfied if purpose is to be an effective sanction of authority. Subordinates must have

confidence that the command is issued in furtherance of a purpose with which they are in sympathy. Second, subordinates must have confidence that the command will be effective in achieving this purpose. This confidence may be based less on their knowledge of the correctness of the command, than on their faith in the ability of those who issued the command; their recognition that those in authority have information they do not have; and their realization that their own efforts will not be effective in reaching the desired objective without some coordination from above.

Within limits, subordinates will even accept commands they know to be incorrect because they do not wish to challenge or unsettle a system of authority that they believe to be beneficial to their aims in the long run. There are, however, restrictions in the authority relationship. In a very real sense, the leader is merely a bus driver whose passengers will leave their leader unless their leader takes them in the direction they wish to go. Thus, subordinates give their leader only minor discretion as to the road to be followed.

Three functions of authority deserve special notice:

- (1) Authority enforces the responsibility of the individual;
- (2) Authority secures expertise in decision-making; and
- (3) Authority permits coordination of activities.

Simon (1947) notes that the core of many of the more important social institutions consists of a system of authority, and a set of sanctions for enforcing the authority relationship. National government is the primary example, but the law of property, the church, and even families are included in this category (see Commons, 1934). Authority refers to the *acceptance* by subordinates of the decisions of the leader, and not the power of the leader to apply sanctions in the case of non-compliance.

Communication. Both Barnard (1938) and Simon (1947) see communication as central to a theory of organization. Simon (1947) argues that without communication there can be no organization, for there is no possibility then for the group to influence the behavior of the individual. Organization members sometimes use informal communication to advance their personal goals. From this informal behavior the phenomenon of cliques arises -- groups that build up an informal network of communications and use this informal network as a means of securing power in the organization. Rivalry among cliques, in turn, may lead to general tensions in social relationships and defeat the purpose of the informal communications system.

Simon (1947) conjectures that weakness of the formal system of communications and failure to secure adequate coordination through that system probably encourages the development of cliques. A great deal of communication goes under the head of 'gossip.' In many organizations the 'grapevine' probably plays, on the whole, a constructive role. Its chief disadvantages are, first, that it discourages frankness, since confidential remarks may be spread about, and second, that the information transmitted by the grapevine is often (deliberately or inadvertently) inaccurate. On the other hand, the grapevine is valuable as a barometer of 'public opinion' in the organization.

It is also important that information does not automatically transmit itself from its point of origin to the rest of the organization; the individuals who first obtain the information must transmit this information. In transmitting the information, organizational members will naturally be aware of the consequences its transmission may have for them. When organizational members believe that the boss is going to be angered by the news, the news is likely to be suppressed. Hence, information tends to be transmitted upward in the organization only if (1) its transmission will not have unpleasant consequences for the transmitters, or (2) the superior will hear of it

anyway from other channels, and it is better to tell the superior first, or (3) it is information that the superior needs in dealings with corporate leaders, and the superior will be displeased if caught without the information.

In addition, there is often failure to transmit information upward simply because subordinates cannot visualize accurately what information their superior needs. A major communication problem, then, of the higher levels of the organization hierarchy is that much of the information relevant to the decisions at this level originates at lower levels, and may not reach the higher levels unless the executive is extraordinarily alert. Simon (1947) also states that there is a converse problem that arises when a superior withholds information from subordinates. This omission, again, may be accidental -- the superior does not realize that subordinates need the information. On the other hand, superiors may use their exclusive possession of information as a means of maintaining authority over subordinates.

In an argument picked up by authors reviewed later in this book, notably Nelson and Winter (1982), Simon (1947) maintains that organizations, to a far greater extent than individuals, need artificial 'memories.' Organizational routines that would become habitual in the case of the individual must be recorded in manuals for the instruction of new organization members. Among the repositories that organizations use are records systems, files, libraries, and follow-up systems. Simon (1947) also observes the importance of motivation: Every effective teacher recognizes that motivation is key to the learning process. Furthermore, personal motives may lead organization members to try to divert the communication system to their own uses, and may lead organization members to withhold information from superiors and colleagues.

The Criterion of Efficiency. Simon (1947) notes that the simplicity of the efficiency criterion in commercial organizations is due, in large part, to the fact that money provides a common denominator for the measurement of both output and income, and permits commercial organizations to be directly compared. Underlying all organizational decisions is a limitation -- a scarcity -- of available resources. This scarcity is the fundamental reason why time and money are costs. Because they are limited in quantity, their application to one organization purpose prevents the realization of alternative possibilities. The criterion of efficiency dictates the choice of alternatives that produces the largest economic result for the given application of resources. Simon (1947) argues that the concept of perfect efficiency is not required. Actual problems, as they present themselves to the decision-maker, are always concerned with *relative* efficiencies, and no measure of *absolute* efficiency is needed. Furthermore, Simon (1947) does not assert that the criterion of efficiency dominates executives' decisions.

The Anatomy of Organization. If there were no limits to human rationality, organizational theory would be barren in Simon's (1947) view. Organization theory would consist of a single precept: Always select the alternative, among those available, that lead to the most complete achievement of desired goals. The need for an organizational theory resides in the fact that there *are* practical limits to human rationality, and that these limits are not static, but depend upon the organizational environment in which the individuals' decisions take place. The task of organization is to design this environment so that individuals approach as close as practicable to rationality (judged in terms of the organization's goals) in their decisions.

In certain situations it is possible to re-orient individuals from identification with a sub-goal of the organization to identification with a broader and more inclusive goal. When a particular item of knowledge is needed repeatedly in decision, the organization can anticipate

this need and, by providing individuals with this knowledge prior to decision, can extend their area of rationality. This knowledge flow is the basic task of organization -- to provide each 'operative' employee with an environment of decision of such a kind that behavior that is rational from the standpoint of this environment is also rational from the standpoint of group values and the group situation. Simon (1947) concludes that the assumption so often made in organizational studies, that an arrangement is effective because it exists, is a circular argument of the worst sort. The only procedure of evaluation that can possibly be valid is the comparison of alternative organization arrangements in terms of their objective results. This procedure is a pragmatic test of what works in practice.

Information Processing. Information need not be processed just because the information is there. Nor should individuals believe that getting more information will always help solve their problems. In some cases, seeking more information indicates a touching faith in more water as an antidote to drowning. Simon (1947) counsels that (social) science does not advance by piling up information -- science organizes information and compresses it. In scientific inquiry, 'knowing' refers to 'knowing parsimoniously.'

Any division of labor among decisional sub-systems creates externalities, which arise out of the interdependencies among the sub-systems that are ignored. What is required for the efficiency of the overall system is a factorization that minimizes these externalities and consequently permits a maximum degree of decentralization of final decisions to the sub-systems, and a maximum use of relatively simple and cheap coordinating devices, like the market mechanism, to relate each of the decisional sub-systems with the other sub-systems.

Simon (1947) argues that the information-processing systems of modern civilization



swim in an exceedingly rich soup of information. In a world of this kind, the scarce resource is not information; it is processing capacity to attend to information. *Attention is the chief bottleneck in organizational activity, and the bottleneck becomes narrower and narrower as we move to the tops of organizations, where parallel processing capacity becomes less easy to provide without damaging the coordinating function that is a prime responsibility of these levels.* Thus, the inherent capacity limits of information-processing systems impose at least two requirements on organizational design: (1) that the totality of decision problems be factored in such a way as to minimize the interdependence of the components, and (2) that the entire system be structured to conserve the scarce resource of attention.

This is a good point to turn to further theoretical developments found in March and Simon (1958) – a book that moves from a ‘closed rational system model’ to an ‘open rational system model’ (Scott, 1987: 100) of the organization. The organization is viewed as evolving toward both increased order and increased complexity. It is a work that provides new insight into the coping mechanisms of the organization.

March, James G., and Herbert A. Simon (1958). Organizations. New York, NY: John Wiley & Sons.

March and Simon (1958) focus on the history of formal organizations. Taking the perspective of social psychologists, March and Simon (1958) are interested in what influences impinge upon individuals from their environment and how individuals respond to such influences. March and Simon (1958) argue that roles in organizations tend to be highly elaborated, relatively stable, and defined to a considerable extent in explicit and even written terms. It is this predictability that enables organizations to deal in a coordinated way with their environments.

March and Simon (1958) take the viewpoint that a decision-maker can be usefully regarded as an information-processor. March and Simon (1958) provide a picture of a choosing, decision-making, problem-solving individual who can do only one or a few things at a time, and who can attend to only a small part of the information recorded in memory and presented by the environment.

March and Simon (1958) note that task allocations are efficient to the extent that such task allocations are based upon similarities in activities that are recognized as yielding important complementarities in task performance. The key idea is to search for complementarities, or in modern terminology, “economies of scope” (Baumol, Panzar and Willig, 1982; Teece, 1980). Beyond this point, solution of the task assignment problem requires empirical knowledge of the specific empirical complementarities that exist.

Behavior in the organization is not determined in advance and once and for all by a detailed blueprint and schedule. Even if it is highly “routinized,” the routine has the character of

a *dynamic capability* rather than a fixed program. March and Simon (1958: 48) provide their own general model of intra-organizational decisions. The essential steps of March and Simon's (1958) behavioral model are:

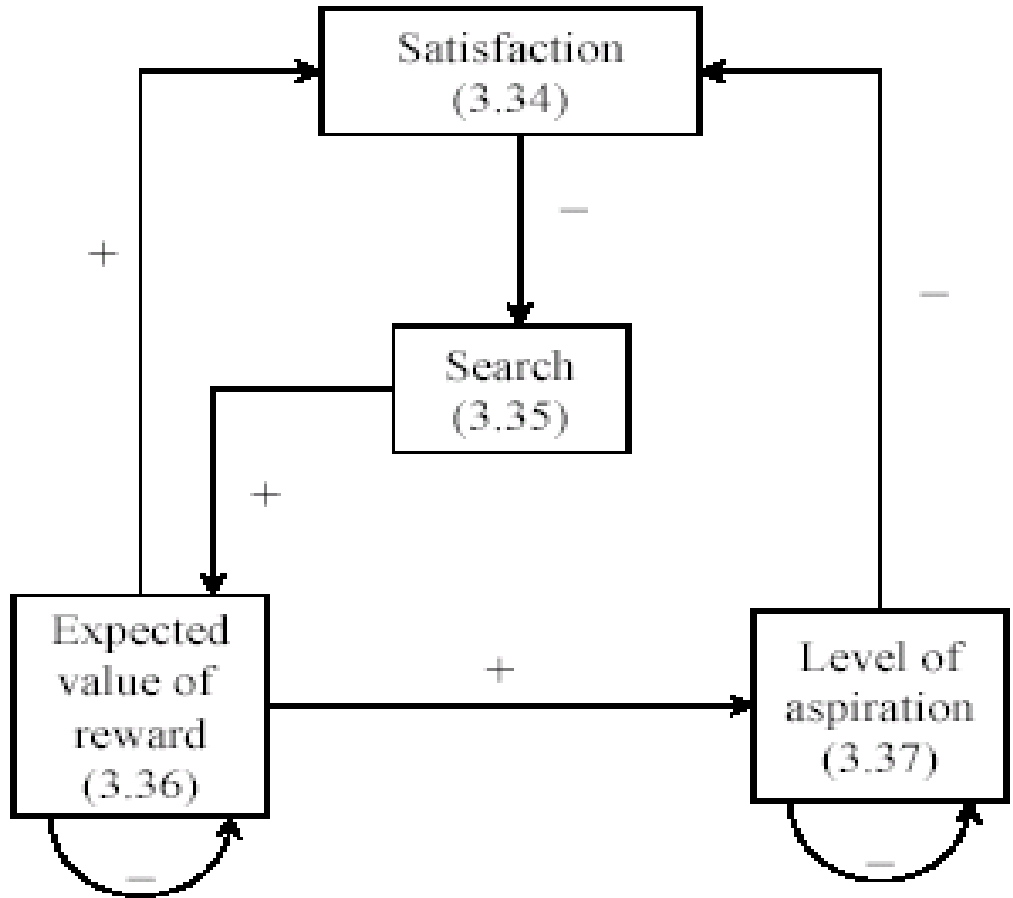
- [1] The lower the *satisfaction* of the individual, the more *search* for alternative programs the individual will undertake;
- [2] The more search, the higher the *expected value of reward*;
- [3] The higher the expected value of the reward, the higher the *satisfaction*;
- [4] The higher the expected value of the reward, the higher the *level of aspiration* of the individual; and
- [5] The higher the level of aspiration, the lower the satisfaction.

-----  
Insert Figure 1 from March and Simon (1958: 49)  
-----

Organizational Rewards. March and Simon (1958) note that many organizational models had historically tended to relegate the reward schemes of management to the background. However, March and Simon (1958) insists that a model of a decision-maker that does not give a prominent place to economic incentives is, for most humans, a poor predictive model. Further, March and Simon (1958) argue that an organization with a promotional scheme that essentially rewards seniority will be less productive than one that relates promotion to some index of productivity.

March and Simon (1958) emphasize that since employees are often cynical regarding announced performance criteria, the factors affecting the subjective operationality of performance standards are important. March and Simon (1958) also observe that, in general, the

**Figure 1. General Model of Adaptive-Motivated Behavior**



Source: March and Simon (1958: 49)

introduction of an incentive wage scheme results in increased production over a straight hourly or day rate, and the introduction of a flat rate payment in place of a former incentive system depresses production. An employment contract based on a flat rate typically is regarded as controlling the *type* of activities performed, but not the *rate* at which activities are performed.

March and Simon (1958) note that the greater the vertical mobility within an organization, the stronger the identification of individuals with the organization. Expectations of vertical mobility create expectations of interactions as well as felt similarities between subordinates and superiors.

Motivational Constraints: The Decision to Participate. The decision to participate is at the core of the theory of what Barnard (1938) and Simon (1947) call "organizational equilibrium:" the conditions of survival of an organization. Equilibrium reflects the organization's success in arranging compensations to its members that are adequate to motivate their continued participation. The Barnard-Simon theory of organizational equilibrium is essentially a theory of motivation – i.e., a statement of the conditions under which an organization can induce its members to continue their participation, and hence increase the likelihood of organizational survival. March and Simon (1958) describes the chief participants of most business organizations, and generally focus on the following five major stakeholders: employees, investors, suppliers, distributors, and consumers. Most obvious in any catalogue of organizational participants are the employees, including the management.

In at least one respect, an employee's relationship to the organization is quite different from that of the other stakeholders. In joining the organization employees accept an authority relationship. Employees agree that within certain limits (defined both explicitly and implicitly by

the terms of employment contracts) the employees will accept as the premises of their behavior instructions supplied to them by the organization.

On the assumption that employees act in a subjectively rational manner, March and Simon (1958) predict the scope of the authority relationship from a knowledge of the inducements and contributions of the employees and other organization members. Employees are willing to enter into employment contracts only if it does not matter to them "very much" what activities (within the zone of acceptance agreed to in the contracts) the organization will instruct them to perform, or if employees are compensated in some way for the possibility that the organization will impose unpleasant activities on them. It is advantageous to subject employees to the organization's authority in those aspects that are of relatively great interest to the employer, comparatively unimportant to the employees, and about which the employer cannot make accurate predictions much in advance of performance.

The problems of both defining and enforcing the "employment contract" are matters of concern, and potential conflict, for all organizational participants. Whether dissatisfaction with the organization leads to withdrawal from the organization depends on whether the participants perceive the "employment contract" as given or as subject to change. Where the contract is viewed as unchangeable, the only options are "accept" or "reject." Where the contract can be changed, participation by no means precludes internal conflicts and bargaining.

Conflict in Organizations by a Bargaining Outcome. March and Simon (1958) argue that game theory, in its original form, was no more satisfactory than neoclassical economic theory in providing an exact prediction of the outcome of a bargaining situation. What game theory offered was a specification of a set of feasible outcomes -- the "solution" of the game. For

example, in the case of highly specialized executives bargaining with their organization over salary, the salary paid will be somewhere between the economic value of the best alternative available to the executives elsewhere (i.e., what the executives can guarantee to themselves without cooperation) and the cost of the organization of hiring and training replacements (i.e., what the organization can guarantee to itself without cooperation). Since this feasible range may be quite wide, the theory is not overly helpful for providing reasonably precise economic predictions. March and Simon (1958) also provide the critical comment that, with rare exceptions, bargaining theory has operated in an empirical vacuum. The assumptions about human motivations and behaviors have usually been made on the basis of introspection, inspection of special cases, and mathematical tractability.

Cognitive Limits on Rationality. How does the rationality of "organizational man" compare with that of neoclassical "economic man" or with the rational man of modern statistical decision theory? The rational decision-makers of economics and statistical decision theory make "optimal" choices in a highly specified and clearly defined environment:

1. When we first encounter them in a decision-making situation, rational decision-makers already have laid out before them the whole set of alternatives from which they will choose their actions. This set of alternatives is simply "given;" the theory does not tell them how this set of alternatives is obtained.
2. To each alternative is attached a set of consequences -- the events that will ensue if that particular alternative is chosen. Here the existing theories fall into three categories:
  - (a) *Certainty*: theories that assume the decision-maker has complete and precise knowledge of the consequences that will follow on each alternative,
  - (b) *Risk*: theories that assume accurate knowledge of a probability distribution of the consequences of each alternative, and

- (c) *Uncertainty*: theories that assume that the consequences of each alternative belong to some subset of all possible consequences, but that the decision-maker cannot assign definite probabilities to the occurrence of particular consequences (see Knight, 1921).
3. At the outset, the decision-maker has a "utility function" or a "preference ordering" that ranks all sets of consequences from the most preferred to the least preferred.
  4. The decision-maker selects the alternative leading to the preferred set of consequences. In the case of *certainty*, the choice is unambiguous. In the case of *risk*, rationality is usually defined as the choice of that alternative for which the expected utility is greatest. Expected utility is defined here as the average, weighted by the probabilities of occurrence, of the utilities attached to all possible consequences. In the case of *uncertainty*, the definition of rationality becomes problematic.

Some Difficulties in the Neoclassical Theory. There are difficulties with this neoclassical model of rational man. In the first place, only in the case of certainty does the neoclassical model agree well with common-sense notions of rationality. In the case of uncertainty, especially, there is little agreement, even among exponents of statistical decision theory, as to the "correct" definition of rationality, or whether, indeed, the term "correct" has any meaning here.

A second difficulty with existing models of rational man is that these models make three exceedingly high demands upon the choice-making mechanism. These models assume (1) that all the alternatives of choice are "given;" (2) that all of the consequences attached to each alternative are known; and (3) that the rational man has a complete utility-ordering for all possible sets of consequences.

Routinized and Problem-Solving Responses. As a challenger to the neoclassical theory of rational choice, the theory of rational choice put forth by March and Simon (1958) incorporates two fundamental characteristics: (1) Choice is always exercised with respect to a



limited, approximate, simplified "model" of the real situation; and (2) The elements of the definition of the situation are not "given," but are themselves the outcome of psychological and sociological processes, including the choosers' own activities, and the activities of others in the choosers' environments.

Activity (individual or organizational) can usually be traced back to environmental stimuli of some sort, e.g., customer orders. The responses to stimuli are of various kinds. At one extreme, a stimulus evokes a response that had been developed and learned at some previous time as an appropriate response for a stimulus of this classification. This response is the "routinized" end of the continuum, where a stimulus calls forth a standard operating procedure almost instantaneously. At the other extreme, a stimulus evokes problem-solving activities directed toward finding performance activities with which to complete the response. Problem-solving activities can generally be identified by the extent to which these activities involve *search*: search aimed at discovering alternatives of action, or consequences of action. "Discovering" alternatives may involve inventing and elaborating whole performance programs where these programs are not already available in the repertory of the problem-solver.

*Search is partly random, but in effective problem solving search is not blind. The design of the search process is itself often an object of rational decision.* Finding the optimal alternative is a different problem from finding a satisfactory alternative. An alternative is *optimal* if: (1) there exists a set of criteria that permits all alternatives to be compared; and (2) the alternative in question is preferred by these criteria to all other alternatives. An alternative is *satisfactory* if: (1) there exists a set of criteria that describes minimally satisfactory alternatives, and (2) the alternative in question meets, or exceeds, all these criteria.

Most human decision-making, whether individual or organizational, is concerned with

the discovery and selection of satisfactory alternatives; only in exceptional cases is it concerned with the discovery and selection of optimal alternatives. To optimize often requires processes several orders of magnitude more complex than processes required to satisfy. An example is the difference between searching a haystack to find the *sharpest* needle in the haystack and searching the haystack to find a needle sharp enough to sew with adequately.

An "optimizing" rule would be to set the standard at the level where the marginal improvement in alternatives obtainable by raising the standard would be just balanced by the marginal cost of searching for alternatives meeting the higher standard. Of course, in practice, the "marginal improvement" and the "marginal cost" are seldom measured in comparable units, or with much accuracy. Thus, the "optimizing" rule is more a *reconstructed logic* used by a researcher in order to make predictions of behavior, rather than characterizing the *logic-in-use* (Kaplan, 1964) of the manager in the actual decision-making process.

Performance Programs. March and Simon (1958) argue that under certain circumstances the search and choice processes are abridged. At the limit, an environmental stimulus may evoke immediately from the organization a highly complex and organized set of responses. Such a set of responses is called a *performance program*. Situations in which a relatively simple stimulus sets off an elaborate program of activity without any apparent interval of search, problem solving, or choice are not rare. Knowledge of the performance program of an organization permits one to predict in considerable detail the behavior of members of the organization.

March and Simon (1958) first argue that organizations attempt to influence employees by specifying standard operating procedures, and attaching organizational rewards and penalties to them. Second, performance programs are important parts of the coordination system in the

organization. These performance programs help fulfill the needs for inter-departmental predictability. Insofar as performance programs are to function as controls, the programs must be linked to variables that are observable and measurable.

March and Simon (1958) expect performance program content to be a function of the *ease of observing job activities*, the *ease of observing job output*, and *the ease of relating activities to output*. Discretion available to the organizational participants is a function of their performance programs and in particular the extent to which the programs specify activities (means), and the extent to which these programs specify products or outcomes (ends).

March and Simon (1958) observe that in organizations there generally is a considerable degree of parallelism between the hierarchical relationships among members of the organization and the hierarchical relations among program elements. That is to say, the programs of members of higher levels of the organization have as their main output the modification or initiation of programs for individuals at lower levels. An important objective of standardization is to widen, as far as possible, the range of situations that can be handled by combination and re-combination of a relatively small number of elementary programs.

March and Simon (1958) contend that rational behavior involves substituting for complex reality a model of reality that is sufficiently simple to be handled by problem-solving processes. In organizations where various aspects of the whole complex problem are being handled by different individuals and different groups of individuals, a fundamental technique for simplifying the problem is to factor the problem into a number of nearly independent parts, so that each organizational unit handles one of these parts and can omit the others from its definition of the situation.

March and Simon (1958) note that the tendency of members of an organizational unit to evaluate action only in terms of sub-goals, even when these sub-goals are in conflict with the goals of the larger organization, is reinforced by at least three cognitive mechanisms. The first cognitive mechanism is located within the individual decision-maker, the second mechanism within the organizational unit, and the third mechanism in the environment of the organizational unit.

In the individual there is cognitive reinforcement through selective perception and rationalization. The propensity of individuals to see things that are consistent with their established frame of reference is well established in organizational psychology. Perceptions that are discordant with the frame of reference are filtered out before they reach consciousness, or are reinterpreted or "rationalized" so as to remove the discrepancy. The frame of reference serves just as much to validate perceptions as the perceptions do to validate the frame of reference.

Within the organization unit, content of in-group communication provides cognitive reinforcement. Such communication affects the focus of information and thereby increases sub-goal persistence. The vast bulk of our knowledge of fact is not gained through direct perception but through the second-hand, third-hand, and nth-hand reports of the perceptions of others, transmitted through the channels of social communication. Two principal types of in-groups are of significance in filtering: in-groups with members in a particular organizational unit, and in-groups with members in a common profession.

Finally, there is reinforcement through selective exposure to environmental stimuli. The division of labor in the organization affects the information that various members receive. This differentiation of information contributes to the differentiation of sub-goals. For example, sales

personnel live in an environment of customers; company treasurers live in an environment of bankers; and each sees a quite distinct part of the business world.

March and Simon (1958) observe that weather persons makes observations of temperature, humidity, and barometric pressure, but may communicate only their conclusions in the form of weather forecasts. In organizational communication, evidence is replaced with conclusions drawn from that evidence, and these conclusions then become the "facts" on which the rest of the organization acts.

When a means of testing actions is perceived to relate to a particular goal or criterion with possible courses of action, the criterion will be called *operational*. Otherwise, the criterion will be called non-operational. For some purposes, we need the further distinction between cases where means-end relations can be evaluated prior to action, and cases where means-end relations can be evaluated only after the fact. March and Simon (1958) call operational goals in the former case operational *ex ante*, and in the latter case, operational *ex post*.

The goal of "promoting the general welfare" is frequently a part of the definition of the situation in governmental policy-making. It is a non-operational goal because this goal does not provide (either *ex ante* or *ex post*) a measuring rod for comparing alternative policies. Strictly speaking, whether a goal is operational or non-operational is not a yes-no question. There are all degrees of "operationality." It will often be convenient, however, to refer simply to the two ends of the continuum.

Important circumstances causing the substitution of sub-goals for more general goals as the criteria for decision-making occur when the sub-goals are perceived as operational, and the goals are perceived as non-operational. For example, a business firm may understand to some degree how its specific actions affect its market share, but may understand less surely how its

actions affect long-term profitability. In such circumstances, the sub-goal of maintaining a particular market share may become the effective criterion of action -- the operational goal.

When a number of individuals are participating in a decision-making process, and these individuals have the same operational goals, differences in opinion about the course of action will typically be resolved by predominantly analytic processes, i.e., by the analysis of the expected consequences of courses of action for realization of the shared goals. When either of the postulated conditions is absent from the situation (when the goals are not shared, or when the shared goals are not operational and the operational sub-goals are not shared), the decision will typically be reached predominately by bargaining processes.

Interdependence does not by itself cause difficulty if the pattern of interdependence is stable and fixed. For in this case, each sub-program can be designed to take account of all the other sub-programs with which it interacts. Difficulties arise only if program execution rests on contingencies that cannot be predicted in advance. In this case, coordinating activity is required to secure agreement about the estimates that will be used as the basis for action, or to provide information to each sub-program unit about the relevant activities of the others. Hence, March and Simon (1958) arrive at the research proposition that the more repetitive and predictable the situation, the greater the *tolerance for interdependence*.

Communication and Coordination. An important method for increasing the organization's tolerance for interdependence is to increase the efficiency of communication by making it possible to communicate large amounts of information with relatively few symbols. An obvious example is the blueprint, which provides a common plan stated in detail. A blueprint uses a carefully defined, highly developed "language" or set of symbolic and verbal conventions.

Because of this standardized language, a blueprint can convey large quantities of information. The same attention to standardization of language is seen in accounting systems and other reporting systems that use numerical data. Accounting definitions and blueprint conventions are examples of a still more general phenomenon: technical languages, whose symbols have definite and common meanings to the members of an organization.

March and Simon (1958) observe that the world tends to be perceived by the organization members in terms of the particular concepts that are reflected in the organization's vocabulary. The particular categories and schemes of classification that the organization employs are reified, and become, for members of the organization, attributes of the world rather than mere conventions.

Organization Structure and the Boundaries of Rationality. March and Simon (1958) maintain that because of the limits of human intellectual capacities in comparison with the complexities of the problems that individuals and organizations typically face, rational behavior calls for simplified models that capture the main features of a problem without capturing all of a problem's complexities. The simplifications have a number of characteristic features:

- Optimizing is replaced by satisficing;
- Alternatives of action and consequences of action are discovered sequentially through search processes;
- Repertories of action programs are developed by organizations and individuals and these repertories serve as the alternatives of choice in recurrent situations;
- Each specific action program deals with a restricted range of situations and a restricted range of consequences; and
- Each action program is capable of being implemented in semi-independence of the others – these action programs are only loosely coupled together.

This "one-thing-at-a-time" or "ceteris paribus" approach to adaptive behavior is fundamental to organization structure. Organization structure consists of those aspects of the pattern of behavior in the organization that are relatively stable and that change only slowly. If behavior in organizations is "intendedly rational," then March and Simon (1958) expect aspects of the behavior to be relatively stable that either (a) represent adaptations to relatively stable elements in the environment, or (b) are the learning programs that govern the process of adaptation.

March and Simon (1958) maintain that a great deal of the inertia of "going concerns" can be explained on the basis of (economic and psychological) sunk costs. A simple example is whether to move to a new location with high moving costs. In addition to tangible sunk costs, persistence comes about primarily because the individual or organization does not search for, or consider, alternatives to the present course of action unless that present course is in some sense "unsatisfactory." March and Simon (1958) also suggest a "Gresham's Law" of planning: Daily routine drives out planning. Stated less cryptically, March and Simon (1958) predict that when an individual is faced both with highly programmed and highly un-programmed tasks, the highly programmed tasks tend to take precedence over the highly un-programmed tasks even in the absence of strong over-all time pressure. Although left unstated by March and Simon (1958), we may anticipate that problems of persistence can be greater for long-term strategy formulation and implementation.

Concluding Comments. Scott (1987) notes that there are important differences between Simon (1947), and March and Simon (1958). While there is still a concern with the cognitive limits of individual decision-makers and with how organizational structures can help to support



improved decision-making, March and Simon (1958) place a greater emphasis on the variable nature of challenges posed by tasks and environments. The organization is viewed as more open to its environment. While some performance programs can be routinized, other performance programs must be problem-solving responses, requiring the decision-maker to exercise more discretion in the face of greater uncertainty. Moreover, it is recognized that some organizations face such volatile environments that these organizations institutionalize innovation, devising programs for “routinely” changing existing program, often rapidly.

In conclusion, my understanding of the evolution of management theory in the 1947-1958 period is that in comparison with Simon (1947), March and Simon (1958) provide a stronger sense that organizations face environments of varying complexity. Furthermore, organizations must adjust their internal decision-making procedures to take these variations into account, and some environments impose levels of complexity that organizations cannot manage unless these organizations impose simplifying restrictions on the information processed.

We turn from March and Simon (1958), to the next landmark in the behavioral theory of the firm by Cyert and March (1963). Cyert and March (1963) focus on a small number of key economic decisions made by the firm and develop process-oriented models of the firm.

Cyert, Richard M., and James G. March (1963). A Behavioral Theory of the Firm. Englewood Cliffs, N.J.: Prentice-Hall.

Cyert and March (1963) are concerned with the business firm and the way the business firm makes economic decisions. Cyert and March (1963) make detailed observations of the processes and procedures by which firms make decisions, and use these observations as a basis for a theory of decision-making in business organizations. Cyert and March (1963) argue that one way to understand modern organizational decision-making is to supplement the microeconomic study of strategic factor markets with an examination of the internal operation of the business firm -- to study the effects of organizational structure and conventional practices on the development of goals, the formation of expectations, and the implementation of choices.

Cyert and March (1963) make four major research commitments:

- To focus on the small number of key economic decisions made by the firm;
- To develop process-oriented models of the firm;
- To link models of the firm as closely as possible to empirical observations; and
- To develop a theory with generality beyond the specific firms studied.

Cyert and March's (1963) conception of the theory-building task is that of constructing a theory that takes the firm as its basic unit; considers the prediction of firm behavior with respect to such decisions as price, output, and resource allocation as its goal; and *emphasizes the actual process of organizational decision-making*. In order to build the behavioral theory of the firm, Cyert and March (1963) develop four major sub-theories concerning:

- Organizational Goals;
- Organizational Expectations;
- Organizational Choice; and
- Organizational Control.

Organizational Goals. A theory of organizational goals considers how goals arise in an organization, how goals change over time, and how the organization attends to these goals. Cyert and March (1963) view an organization as a coalition of stakeholders, with some of these stakeholders organized into sub-coalitions. In a business organization the coalition members include managers, workers, stockholders, suppliers, customers, lawyers, tax collectors, regulatory agencies, and so on. Clearly then, organizational goals must deal successfully with the potential for internal goal conflicts inherent in a coalition of diverse individuals and groups.

Since the existence of unresolved conflicts among organizational stakeholders is a key feature of organizations, it is difficult to construct a useful descriptively accurate theory of the organizational decision-making process if we insist on internal goal consistency. Cyert and March (1963) do not insist then that such a decision-making process necessarily produces consistent organizational goals.

Since individuals have limited capacities, and limited time, to devote to any particular aspect of the organizational system, such limitations constrain the bargaining process. As an adaptive response, coalition members are motivated to develop mutual control systems such as the budget, and the allocation of tasks by the division of labor and specialization. A budget becomes a precedent for future budgets; an allocation of tasks becomes a precedent for future task allocations. Thus, coalition agreements are institutionalized into semi-permanent arrangements.

Another important mechanism for dealing with stakeholder conflicts is the sequential attention to conflicting goals. A consequence of this mechanism is that organizations ignore many conditions that outside observers see as direct contradictions. The decentralization of decision-making (and goal attention), the sequential attention to goals, and the adjustment in

organizational slack, which acts as a cushion in down times, permit the business firm to make decisions with inconsistent goals under many (and perhaps most) conditions.

Organizational Expectations. A theory of organizational expectations considers how and when an organization searches for information or new alternatives, and how information is processed through the organization. Expectations are by no means independent of hopes, wishes, and the internal bargaining needs of sub-units in the organization. Information about the consequences of specific courses of action in a business organization is frequently hard to obtain and of uncertain reliability. As a result, both conscious and unconscious biases in expectations are introduced. Thus, local priorities and perceptions obtain. In addition, there is some evidence of more conscious manipulation of expectations. Communication in a complex organization includes considerable biasing and influence activities, but also considerable bias correction as well. In addition, organizations often protect themselves from the worst effects of influence activities by focusing on verified data in lieu of uncertain estimates, and by using easily checked feedback information.

Organizational Choice. A theory of organizational choice needs to characterize the process by which the alternatives available to the organization are ordered and selected. Organizational decisions depend on information estimates, and expectations that ordinarily differ appreciably from reality. These organizational perceptions are influenced by some characteristics of the organization and its procedures. Second, organizations consider only a limited number of decision alternatives. Finally, organizations vary with respect to the amount of resources that such organizations devote to their organizational goals on the one hand, and sub-organizational and individual goals on the other hand. The firm is considered to be an adaptively rational

system in which the firm learns from experience. General choice procedures are summarized in terms of three basic principles:

- *Avoid Uncertainty.* The firm looks for procedures that minimize the need for predicting uncertain future events. One method uses short-run feedback as a trigger to achieve action, another accepts (and enforces) standardized decision rules.
- *Maintain the Rules.* Once the firm has determined a feasible set of decision procedures, the organization abandons them only under duress.
- *Simplify the Rules.* The firm relies on individual "judgment" to provide flexibility around simple rules.

Organizational Control. A theory of organizational control specifies the difference between executive choice in an organization, and the decisions actually implemented. Organizational control within an organization depends on the elaboration of standard operating procedures. It is hard to see how a theory of the firm can ignore the effect of such organizational procedures on decision-making behavior within the organization. The effects fall into at least four major categories: (a) Effects on individual goals within the organization; (b) Effects on individual perceptions of the environment; (c) Effects on the range of alternatives considered; and (d) Effects on the managerial decision rules used. Cyert and March's (1963) basic theory of organizational control assumes:

- Multiple, changing, acceptable-level goals. The criterion of choice is that the alternative selected meets the demands (goals) of the coalition;
- An approximate sequential consideration of alternatives. The first satisfactory alternative evoked is accepted. When failure occurs, search is intensified; and
- The organization seeks to avoid uncertainty by following standard operating procedures and a policy of reacting to feedback rather than forecasting the environment.

Summary of Cyert and March's (1963): *A Behavioral Theory of the Firm*. Cyert and March (1963) propose two major organizing devices: (1) A set of variable concepts; and (2) A set of relational concepts. The variable concepts discussed above are organizational goals, organizational expectations, organizational choice, and organizational control. There are also four major relational concepts:

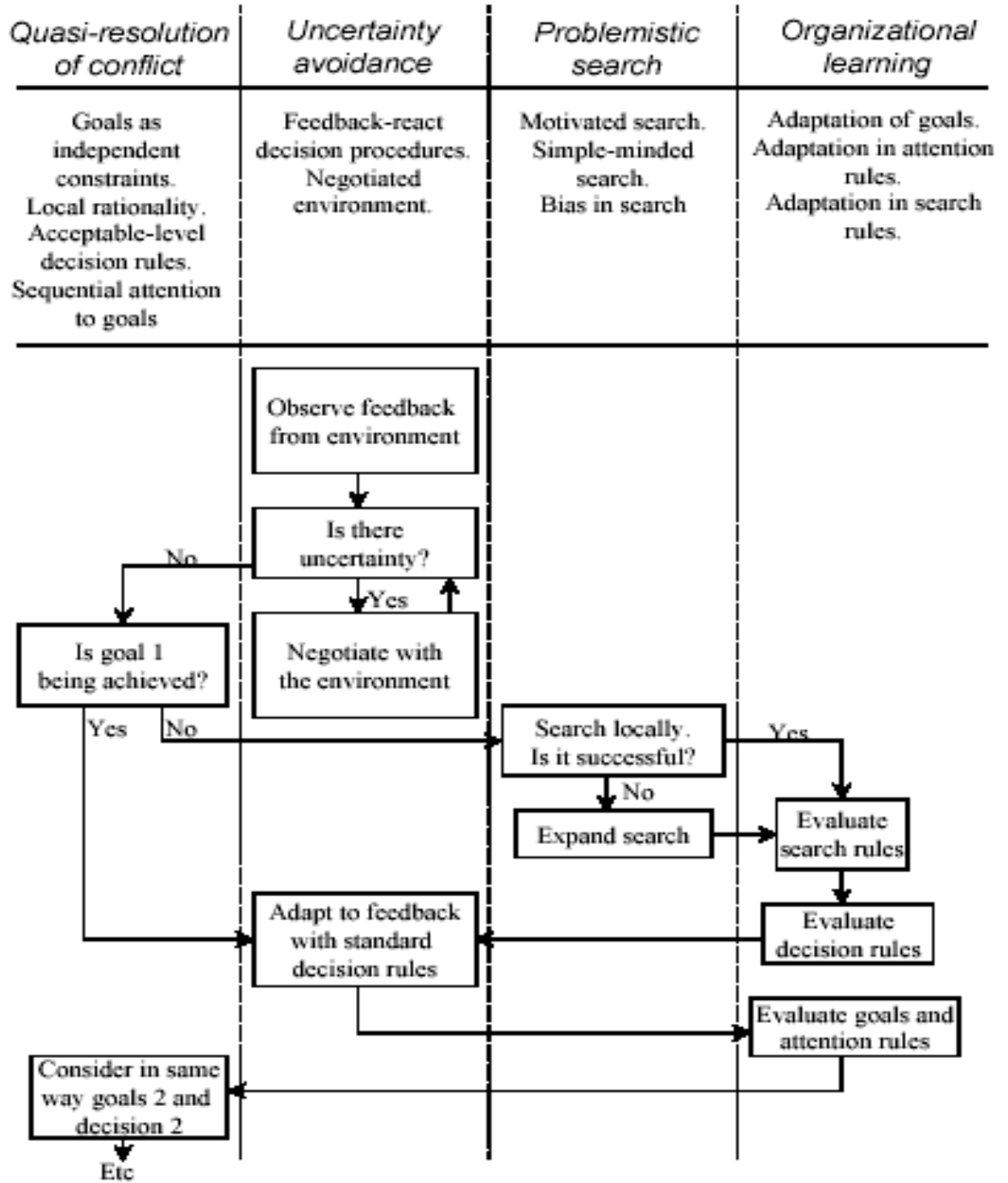
- Quasi Resolution of Conflict;
- Uncertainty Avoidance;
- Problemistic Search; and
- Organizational Learning.

-----  
Insert Figure 2 from Cyert and March (1963: 126)  
-----

Quasi Resolution of Conflict. In keeping with numerous theories of organizations, Cyert and March (1963) assume that the coalition in an organization is a coalition of members having different personal goals. Members require some procedure for resolving conflicts such as acceptable-level decision rules and/or sequential attention to goals.

Uncertainty Avoidance. Cyert and March (1963) submit that organizations typically try to avoid uncertainty: First, organizations avoid the requirement that they correctly anticipate events in the distant future by using decision rules emphasizing short-run reactions to short-run feedback, rather than anticipation of long-run uncertain events. Second, organizations avoid the requirement that they anticipate future reactions of other parts of their environment by arranging a negotiated environment. Organizations impose plans, standard operating procedures, industry tradition, and uncertainty-absorbing contracts on that environment.

Figure 2. Organizational Decision Process in Abstract Form



Source: Cyert and March (1963: 126)

Problemistic Search. Cyert and March's (1963) behavioral models assume that search, like decision-making, is problem-directed. Problemistic search means search that is stimulated by a problem (usually a rather specific one), and is directed toward finding a solution to that problem. Such organizational search is assumed to be motivated, simple-minded, and biased. This bias may reflect training or experience of various parts of the organization. This bias may reflect the interaction of hopes and expectations, and communication biases are expected to reflect unresolved conflicts within the organization.

Organizational Learning. Organizations learn: to assume that organizations go through exactly the same processes as individuals go through seems unnecessarily naive, but organizations exhibit (as do other social institutions) adaptive behavior over time. Cyert and March (1963) focus on adaptation with respect to three different phases of the decision process: adaptation of goals, adaptation in attention rules, and adaptation in search rules. Cyert and March (1963) submit that organizations change their goals, shift their attention, and revise their procedures for search as a function of their experience.

Scott (1987) notes that Cyert and March's (1963) concept of *coalitions* offers the following features:

- The problem of reification is avoided; individuals and groups have interests, and the processes by which these preferences come to be imposed on the organization are specified;
- It is recognized that although individuals and groups are allowed to specify the goals of the organization; there is no presumption that they do so on an equal footing, nor is it assumed that they hold common objectives;
- It is recognized that although individuals and groups impose goals on the organization, in most cases no single individual or group is powerful enough to determine completely the organization's goals; hence, the organization's goals are typically distinct from those of any of its participants;



- Allowance is made for differences in interests among participants. Some, but not all, of these differences may be resolved by negotiation, so at any time conflicting goals may be present; and
- It is recognized that the size and composition of the dominant coalition may vary from one organization to another and within the organization from time to time.

In my judgment, Cyert and March (1963) provides a more conceptually refined and systematic outline of the behavioral theory of the firm that improves upon March and Simon (1958). Three statements can summarize Cyert and March's (1963) arguments: (1) The business firm is a relevant unit of investigation; (2) It is possible to construct a theory of decision-making behavior within such a unit; and (3) Such a theory must focus explicitly on actual organizational decision processes. To this purpose, Cyert and March (1963) show how to construct behavioral models of firm-level decision-making and indicate the basic theoretical framework within which such models are embedded. Cyert and March's (1963) behavioral theory of the firm can be applied to: (a) price and output decisions; (b) internal resource allocations; (c) innovations; (d) competitive dynamics; and (e) predicting the behavior of other organizations.

I hold the hope that current students studying the economics of organization may build upon Cyert and March (1963) and connect this research agenda with other branches of organizational economics, which we consider in subsequent chapters. It is my view that building a science of organization that suppresses issues of bounded rationality and limited information processing by organizational members would lead the strategic management field up a blind alley.

Now that we have examined the classic work of Barnard (1938), and of the "Carnegie School" of Simon (1947), March and Simon (1958), and Cyert and March (1963), we apply the Carnegie framework to the Cuban Missile Crisis, where for thirteen days the United States and the Soviet Union paused at the nuclear precipice (Allison, 1971).

---

## **Application: Explaining the Cuban Missile Crisis**

Based on materials from Graham T. Allison's (1971), The Essence of Decision.

The Cuban missile crisis was a seminal event in the history of the United States. On the days between October 16 and October 28, 1963, the United States and the Soviet Union came dangerously close to nuclear war. Using the Carnegie School framework for explaining the crises, we consider the following question: Why did the Soviet Union place strategic offensive missiles in Cuba?

Why did the Soviet Union place strategic offensive missiles in Cuba? From the Carnegie School framework, explanation for this action requires an identification of the relevant Soviet organizations and displays the patterns of organizational behavior from which the actions emerged. An explanation, from this perspective, must identify trends that reflect established organizations and their somewhat rigid operating procedures and programs. That is, governmental behavior is understood less as deliberate choices (as if from a unitary actor), and more as *outputs* of large organizations functioning according to standard patterns of behavior.

While the final decision to put missiles in Cuba must have been made in the Presidium, the details of this operation – that is the path from the general decision to the actual appearance of operation missiles in Cuba – were probably delegated to appropriate Soviet organizations --- such as the GRU (Soviet military intelligence), the KGB (the Communist party security agency), the SAM (the Soviet Air Defense Command), and a quite separate Soviet military service, the Strategic Rocket Forces. Standard Soviet operations, particularly when nuclear weapons were involved, imposed a very high level of secrecy. Thus, each organization's tendency to follow standard operating procedure was reinforced by a lack of information about the activity of other organizations and the impossibility of an overview of the whole operation. Allison (1971) discusses several instances where contradictory behaviors and anomalies (from the perspective

of a unitary actor model) are explained from the Carnegie School framework. Many crucial details of implementation followed from organizational routines rather than from central choice.

The lesson, as Allison (1971) suggests, is that nuclear crises between machines as large as the United States and the Soviet Union have elements of genuine uncertainty. The information and estimates available to leaders about the situation will reflect organizational goals and routines as well as the facts. The alternatives presented to leaders will be much narrower than the menu of options that would typically be more desirable. The implementation of choices will exhibit unavoidable rigidities of organization's standard operating procedures. In a crisis, the overwhelming problem will be that of control and coordination of large organizations.

---

Finally, we conclude this first chapter with a summary of Simon's (1982) work in the research area of behavioral economics, which is worthy of careful attention by students studying the economics of organization. Simon (1982), leading by example, shows how fruitful social science research can be for those who are not intimidated by disciplinary boundaries and that anything that can improve our understanding of complex organizations should be valued.

Simon, Herbert A. (1982). Models of Bounded Rationality: Behavioral Economics and Business Organization. Cambridge, Mass.: MIT Press.

Simon (1982) argues that organization theory, economics (especially the theory of the firm), and cognitive psychology are all basically concerned with the same phenomena. All three are theories of human decision-making and problem-solving processes; yet each of the three domains has developed in relative isolation from the other two domains. Simon (1982) is concerned with both the causes for this isolation and its remedies.

The Business Firm as an Organization. Simon (1982) notes that the firm of neoclassical economic theory is little more than an entrepreneur who is attached a cost curve or a production function. Since profit maximization and internal efficiency are assumed, there is little room in the neoclassical theory for the familiar institutional characteristics of real business firms. For example, for the facts that one of a business firm's principal inputs typically is labor, a "commodity" that is contracted for on quite a different basis from other commodities, and that decisions are reached within a hierarchy of authority relationships among the employees.

A Formal Theory of the Employment Relationship. Simon (1982) observes that neoclassical economic theory abstracts away the distinctive characteristics of the employment contract, and neoclassical economic theory ignores the most significant features of the organizational process, i.e., the process of actually managing the factors of production, including the input of labor. Simon (1982) sets forth a theory of the employment relationship that re-introduces some of the more important of these empirical realities into the economic model. Perhaps in this way a bridge can be constructed in the discipline of strategic management between economists, with their theories of the firm and of factor allocations, and organization

theorists, with their theories of organization --- a bridge wide enough to permit some free trade of ideas between two intellectual domains that have been isolated from each other.

The authority relationship that exists between an employer and an employee, an economic relationship created by the employment contract, plays a central role in Simon's (1982) theory. Let employer B (for "boss") hire employee W (for "worker"). We say that B exercises authority over W if W permits B to select behavior  $x$ . That is, W accepts authority when W's behavior is determined by B's decision. In general, W will accept authority if the decision is within W's "zone of acceptance."

We say that W enters into an employment contract with B when the former agrees to accept the authority of the latter and the latter, agrees to pay the former a stated wage ( $w$ ). This contract differs fundamentally from a sales contract -- the kind of contract that is assumed in typical formulations of neoclassical price theory. In the sales contract, each contractual party promises a specific consideration in return for the consideration promised by the other contractual party. The buyer (like B) promises to pay a stated sum of money; but the seller (unlike W) promises in return a specified quantity of a completely specified commodity.

W will be willing to enter an employment contract with B only if it does not matter to W "very much" which  $x$  (within the agreed-upon area of acceptance) B will choose or if W is compensated in some way for the possibility that B will choose an  $x$  that is not desired by W (e.g., that B will ask W to perform an unpleasant task).

It will be advantageous to B to offer W added compensation for entering into an employment contract if B is unable to predict with certainty, at the time the contract is made, which  $x$  will be the optimal one, from W's standpoint. That is, B will pay for the privilege of postponing, until some time after the contract is made, the selection of  $x$ . This option to wait has

real options value, which is explained more fully, and illustrated by a mathematical example, in chapter 5.

Simon's (1982) model deals with a particular problem of planning under uncertainty. It analyzes a business situation in which it may be advantageous to postpone a decision (selection of  $x$ ) in order to gain from information obtained subsequently. The postponement of choice may be regarded as a kind of "liquidity preference" where the liquid resource is the employees' time, instead of money.

A Comparison of Organization Theories. According to Simon (1982), the economic theory of the firm and the organization theory of the firm are both concerned with the behavior of a person, or people, trying to achieve certain goals by the manipulation of (strategic) variables at their disposal. The problem of "optimal," "rational," or "efficient" behavior with respect to these goals can be formulated as a problem of finding the maximum (with respect to the strategic variables) of some function that is taken as a measure of success in attaining these goals (e.g., in the theory of the firm, finding the output that maximizes economic profit). Theories of organization, to a greater extent than the economic theory of the firm, have been concerned not only with optimal solutions, but with the whole set of viable solutions -- that is, solutions that permit the survival of the organization (e.g., in the theory of the firm, outputs that yield at least a competitive rate of economic return).

In the neoclassical economic theory of the firm, a single participant, the entrepreneur, is explicitly treated as a rational individual. The other participants -- employees, customers, and suppliers -- enter into the neoclassical theory only implicitly as "conditions" to which the entrepreneur adjusts in finding an economic solution that is optimal to the entrepreneur. One

such condition is the price of the input factor "labor," another is the demand schedule, which describes the behaviors of customers.

In the organizational theory, the participants are generally treated in a more symmetrical way (Barnard, 1938; Simon, 1947). Participants are offered inducements for their participation in the organization. Through their participation, organizational members make contributions to the organization. The organization transforms its members' contributions into inducements that the organization, in turn, distributes to these members.

As a simple example, consider an organization with an entrepreneur, one employee, and one customer. The system of inducements and contributions may then be represented thus:

<b>Participant</b>	<b>Inducements</b>	<b>Contributions</b>
Entrepreneur	Revenue from Sales	Cost of Production
Employee	Wage	Labor
Customer	Goods	Purchase Price

Organization theory has generally been concerned not so much with optimality as with the conditions necessary for organizational survival, that is, the conditions under which the participants will continue to participate. The picture of the firm that is emerging from this research is that of a searching, information-processing, satisficing, allocating mechanism. Human thinking is an important — indeed, the most important — economic resource. The bulk of the productive wealth of our economy is not embodied in factories and machines, but it is to be found in the knowledge and skills stored in individuals' minds.

The Scarcity of Attention. A rabbit-rich world is a lettuce-poor world, and vice versa. Similarly, in an information-rich world, an abundance of information means a dearth of something else: a scarcity of whatever information consumes. Information consumes the attention of its recipients. In an information-rich world, most of the economic costs of

information are the costs incurred by the recipients. It is not enough to know how much it costs to produce and transmit information: we must also know how much it costs, in terms of scarce attention, to receive information. As we shall see later, in chapter 5, this idea informs Penrose's (1959) resource-based theory of managerial attention as the scarce resource, which is the binding constraint on the rate of the growth of the firm.

Many proposals for eliminating information overload (another phrase to describe life in an information-rich world) call for a new information system. An information-processing sub-system (a computer or new organization unit) will reduce the net demand on the rest of the organization's attention only if this sub-system absorbs more information previously received by others than it produces -- that is, if this sub-system listens and thinks more than it speaks. The proper aim of a management information system is not to bring managers all the information they need, but to reorganize the managers' environment of information so as to reduce the amount of time they must devote to receiving information. Restating the organization problem this way leads to a very different system design.

Simon (1982) notes that if a library has only one copy of each book, the library still has a high degree of informational overlap. Simon (1982) conjectures that if half the titles in the Library of Congress were destroyed at random, little of the world's knowledge would be lost. The most important form of redundancy derives from the world's being highly lawful. Facts are lawful if they can be predicted from other facts. We need store only the fraction needed to predict these other facts. This process is exactly what science does: the process of replacing large amounts of unordered facts with tidy statements of orderly relations from which these facts can be inferred.



Simon (1982) observes that the dream of thinking everything out before we act, of making certain we have all the facts, and that we know all the consequences, is a sick Hamlet's dream. It is a dream of someone with no analytical appreciation of the seamless web of causation, the limits of human thinking, or the scarcity of human attention.

Theories of Decision-Making in Economics and Behavioral Science. Simon (1982) notes that the neoclassical economic theory of markets with perfect competition and rational agents is deductive theory that requires almost no contact with empirical data once the assumptions are accepted. Undoubtedly, there is an area of human behavior that fits these assumptions to a reasonable approximation, where the neoclassical microeconomic theory with its assumptions of rationality is a powerful and useful tool. Without denying the existence of this area, or its importance, Simon (1982) observes that neoclassical microeconomic theory fails to include some of the central problems of conflict and dynamics with which organization theory and strategic management have become increasingly concerned. A metaphor helps to show the reason for this failure (Simon, 1982: 255):

Suppose we were pouring some viscous liquid -- molasses -- into a bowl of very irregular shape. What would we need in order to make a theory of the form the molasses would take in the bowl? How much would we have to know about the properties of molasses to predict its behavior under the circumstances? If the bowl were held motionless, and if we wanted only to predict behavior in equilibrium, we would have to know little, indeed, about molasses. The single essential assumption would be that the molasses, under the forces of gravity, would minimize the height of its center of gravity. With this assumption, which would apply as well to any other liquid, and a complete knowledge of the environment -- in this case the shape of the bowl -- the equilibrium is completely determined. Just so, the equilibrium behavior of a perfectly adapting organism depends only on its goal and its environment; it is otherwise completely independent of the internal properties of the organism. If the bowl into which we were pouring the molasses were jiggled rapidly, or if we wanted to know about the behavior before equilibrium was reached, prediction would require much more information. It would require, in particular, more information about the properties of molasses: its viscosity, the rapidity with which it "adapted" itself to the containing vessel and moved toward its "goal" of lowering its center of gravity. Likewise, to predict the short-run behavior of an adaptive

organism, or its behavior in a complex and rapidly changing environment, it is not enough to know its goals. We must know also a great deal about its internal structure and particularly its mechanisms of adaptation.

Simon (1982) argues that broadening the definition of rationality to encompass goal conflicts and uncertainty made it difficult to ignore the distinction between the objective environment in which economic actors "really" live, and the subjective environment that they perceive and to which they respond. When this distinction is made, we can no longer predict their behavior -- even if they behave rationally -- from the characteristics of the objective environment. *We also need to know something about their perceptual and cognitive processes.*

Simon (1982) maintains that models of satisficing behavior are richer than models of maximizing behavior, because models of satisficing behavior consider not only of equilibrium, but also of the method of reaching equilibrium as well. Neoclassical economic theory is a theory of an individual choosing among fixed and known alternatives, to each of which the known consequences are attached. But when perception and cognition intervene between the decision-maker and an objective environment, neoclassical economics no longer proves adequate. We need a description of the choice process that recognizes that alternatives are not given but must be sought, and a description that takes into account the arduous task of determining what consequences will follow from each alternative.

Decision-makers' information about their environment is actually much less than an approximation to the real environment. The term "approximation" implies that the subjective world of decision-makers resembles the external environment closely, but lacks, perhaps, some fineness of detail. In actual fact, the perceived world is quite different from the "real" world. The differences involve both omissions and distortions, and arise in both perception and

inference. The sins of omission in perception are arguably more important than the sins of commission. Decision-makers' mental models of the world encompass only a minute fraction of all the relevant characteristics of the real environment, and these inferences extract only a small fraction of all the information that is present.

Perception is sometimes referred to as a "filter." This term is as misleading as "approximation," and for the same reason: perception implies that what comes through into the central nervous system is really quite a bit like what is "out there." In fact, the filtering is not merely a passive selection of some part of a presented whole, but is an active process involving attention to a very small part of the whole and exclusion, from the outset, of almost all that is not within the scope of attention.

Simon (1982) argues that every human lives in an environment that generates millions of bits of new information each second, but the bottleneck of the perceptual apparatus certainly does not admit more than 1,000 bits per second, and probably much less. Equally significant omissions occur in the processing that takes place when information reaches the brain. There are hosts of inferences that might be drawn from the information stored in the brain that are not in fact drawn. The consequences implied by information in the memory become known only through active information processing, and hence through active selection of particular problem-solving paths from the myriad problem-solving paths that might have been followed.

Theories of Bounded Rationality. Simon (1982) argues that rationality, as is typically defined in the social sciences, denotes behavior that is appropriate to the achievement of given goals, within the limits imposed by given constraints. Those theories that postulate important constraints arising from the limitations of the actors themselves as information processors may be called theories of bounded rationality.

In some sense, chess is a “trivial” game: if the complete decision tree of possible games was fully known, there would be nothing of interest left to play. Unfortunately, the triviality of chess, as viewed from this high level of abstraction, offers no practical guide to a player in actually choosing a move. The proof that guarantees the validity of one (and only one) of three alternatives that the game must have a value of win, lose or draw for White, gives no practically usable method to determine the true outcome. This relative, human difficulty necessitates the use of those incomplete, heuristic methods of playing, which constitute "good" Chess; and without this human difficulty there would be no element of "struggle" and "surprise" in this game.

Simon (1982) emphasizes that the chess player's difficulty in behaving rationally has nothing to do with uncertainty -- whether of consequences or alternatives -- but it is a matter of *complexity*. For there is no risk or uncertainty, in the sense in which those concepts are used in economics or in statistical decision theory, in the game of chess. It is a game of perfect information. No probabilities of future events need enter the calculations, and no contingencies, in a statistical sense, arise.

What we refer to as "uncertainty" in chess is uncertainty introduced into a perfectly certain environment by inability -- computational inability -- to ascertain the structure of that environment. But the result of the uncertainty, whatever its source, is the same; approximation must replace exactness in reaching a decision.

A satisficing decision procedure can often be turned into a procedure for optimizing by introducing a rule for optimal amount of search, or, what amounts to the same thing, a rule for fixing the aspiration level optimally. Thus, the aspiration level in chess might be adjusted,

dynamically, to such a level that the expected improvement in the move chosen, per minute of additional search, would just balance the incremental cost of the search.

Although such a "reconstructed logic" (Kaplan, 1954) is formally possible, to carry it out in practice requires additional information and assumptions beyond those needed for satisficing. First, the values of alternatives must be measured in units comparable with the units for measuring search costs, in order to permit comparison at the margins. Second, the marginal productivity of search -- the expected increase in the value per unit of search time -- must be estimated on some basis. If one were designing a chess-playing program, it is doubtful whether effort spent in attempting to place the program in such a dynamic optimizing framework would be nearly as worthwhile as an equivalent effort given to improving the selectivity of the program's move-generating and move-evaluating heuristics. Research on satisficing procedures has focused primarily on the efficiency of search -- on the nature of the heuristic methods.

Simon (1982) observes that most of the formal techniques that constitute the technical backbone of management science and operations research are procedures for finding the best of a set of alternatives in terms of some criterion. Linear programming and dynamic programming are among the more powerful of these techniques. The dominant approach to problems in this sphere has been to simplify the real-world problems to the point where the formal optimizing models can be used as approximations.

Perhaps the technique most widely used in management science to deal with situations too complex for the application of known optimization methods is simulation. In simulation, the trial and error is supplied by the human investigators rather than by the technique of analysis itself. The satisficing approach has been most often employed in models where "heuristic" or trial-and-error methods are used to aid the search for plausible alternatives. These computational

tools make substantially more tractable the task of matching bounded capabilities with the difficulty of the problems.

From Substantive to Procedural Rationality. Simon (1982) uses the phrase "substantive rationality" to refer to the concept of rationality that developed within economics, and "procedural rationality" to refer to the concept that developed within psychology. Behavior is substantively rational when such behavior is appropriate to the achievement of given goals within the limits imposed by given constraints. Notice that, by this definition, the rationality of behavior depends upon the actors in only a single respect -- their goals. Given these goals, rational behavior is determined entirely by the characteristics of the environment in which such behavior takes place.

Neoclassical economic analysis rests on at least two fundamental assumptions. The first assumption is that economic actors have particular goals, for example, utility maximization or profit maximization. The second assumption is that economic actors are substantively rational. Given these two assumptions, and given a description of a particular economic environment, economic analysis (descriptive or normative) could usually be carried out using such standard tools as differential calculus, linear programming, or dynamic programming.

Thus, the assumption of utility or profit maximization on the one hand, and the assumption of substantive rationality, on the other hand, freed economics from any dependence upon psychology. As long as these assumptions went unchallenged, there was no reason why economists should acquaint themselves with the psychological literature on human cognitive processes or human choice. There was absolutely no point at which the empirical findings of psychological research could be injected into the process of economic analysis. The irrelevance

of psychology to neoclassical economics was complete.

Behavior is procedurally rational when such behavior is the outcome of appropriate deliberation. Its procedural rationality depends upon the process that generated it. Historically, there have been three main categories of psychological research on cognitive processes: learning, problem solving, and concept attainment.

*The search for computational efficiency is a search for procedural rationality, and computational mathematics is a normative theory of procedural rationality.* In this normative theory, there is no point in prescribing a particular substantively rational solution if there exists no procedure for finding that solution with an acceptable amount of computing effort. So, for example, although there exist optimal (substantively rational) solutions for combinatorial problems of the traveling-salesman type, and although these solutions can be discovered by a finite enumeration of alternatives, actual computation of the optimum is infeasible for problems of any size and complexity. The combinatorial explosion of such problems simply outraces the capacities of computers, present and prospective.

*Hence, a theory of rationality for problems like the traveling-salesman problem is not a theory of best solutions -- of substantive rationality -- but a theory of efficient computational procedures to find good solutions -- a theory of procedural rationality.* Notice that this change in perspective involves not only a shift from the substantive to the procedural, but a shift also from concern for optimal solutions to a concern for good solutions (e.g., good decision rules for inventory and work-force smoothing).

Simon (1982) argues that the demands of computability led to two kinds of deviation from neoclassical optimization: simplification of the model to make computation of an "optimum" feasible, or alternatively, searching for satisfactory, rather than optimal choices.

Simon (1982) regards both of these solutions as instances of satisficing behavior rather than optimization. To be sure, using reconstructed logic we can formally view these as optimizing procedures by introducing, for example, a cost of computation and a marginal return from computation, and using these quantities to compute the optimal stopping-point for the computation. But the important difference between the satisficing procedures and the optimizing procedures remain. The problem has been shifted from one of characterizing the substantively optimal solution to one of devising practicable computation procedures for making reasonable choices.

Ignorance of the future prevents decision-makers from behaving in a substantively rational manner; decision-makers can only adopt a rational choice procedure, including a rational procedure for forecasting or otherwise adapting to the future. Once we become interested in the procedures -- the rational processes -- that economic actors use to cope with uncertainty, we must broaden our horizons further. Uncertainty calls forth a whole range of actions. These actions are at least of four kinds:

- Intelligence actions to improve the data on which forecasts are based, to obtain new data, and to improve the forecasting models;
- Actions to buffer the effects of forecasting errors; holding inventories, insuring, and hedging, for example;
- Actions to reduce the sensitivity of outcomes to the behavior of competitors: steps to increase product and market differentiation, for example; and
- Actions to enlarge the range of alternatives whenever the perceived alternatives involve high risk.



As organizational economics and strategic management become more concerned with procedural rationality, they will necessarily have to borrow from psychology, or build for itself a far more complete theory of human cognitive processes. Even if our research interest in strategic management is in normative rather than descriptive behavior, we will need such a theory. There are still many areas of decision -- particularly those that are ill-structured -- where human cognitive processes are more effective than the best available optimization techniques or artificial intelligence methods. A great deal can still be learned about effective decision procedures by studying how humans make choices.

We can expect substantive rationality only in those situations that are sufficiently simple as to be transparent to the decision-maker's mind. In all other situations, we must expect that the decision-maker's mind will use such imperfect information as it has, will simplify and represent the situation as it can, and will make such calculations as are within its powers (Duhaime and Schwenk, 1985). We cannot expect to predict what the decision-maker's mind will do in such situations unless we know what information it has, what forms of representations it prefers, and what algorithms are available to it.

In my judgment, there seems to be no escape from psychology. If organizational economics and strategic management are to deal with uncertainty, they will have to understand how humans in fact behave in the face of uncertainty, and by what limits of information and computability humans are bound. Bobby Fischer, in 1972, played chess differently from Paul Morphy, in 1861. Much of that difference was the result of the knowledge of the game that had cumulated over the century through the collective experience of the whole society of professional chess players. Organizational economics and strategic management are, like chess, inevitably culture-bounded and history-bounded. A business firm equipped with the tools of

operations research does not make the same decisions, for example, concerning inventory management, as it did before it possessed such tools.

*Simon (1982) maintains that (organizational) economics is one of the sciences of the artificial. Organizational economics is a description and explanation of human institutions, whose theory is no more likely to remain invariant over time than the theory of bridge design. Decision processes, like all other aspects of economic institutions, exist inside human heads. Decision processes are subject to change with every change in what humans know, and with every change in their means of calculation. Simon (1982) submits that for this reason the attempt to predict, and prescribe, human economic behavior by deductive inference from a small set of unchallengeable premises must fail, and has failed.*

Simon (1982) suggests that organizational economics will progress as we deepen our understanding of human thought processes; and organizational economics will change as human individuals and human societies use progressively sharpened tools of thought in making their decisions and designing their institutions. A body of theory for procedural rationality is consistent with a business world in which humans continue to think and continue to invent; a theory of substantive rationality is not.

Simon (1982) notes that the shift from theories of substantive rationality to theories of procedural rationality requires a basic shift in scientific style, from an emphasis on deductive reasoning within a tight system of axioms to an emphasis on detailed empirical exploration of complex algorithms of thought. As organizational economics becomes more involved in the study of uncertainty, and more concerned with the complexity of business decision-making, the shift in research program becomes inevitable. Wider and wider areas of organizational

economics and strategic management will replace the over-simplified assumptions of situationally constrained omniscient decision-makers with a realistic (and psychological) characterization of the limits on decision-makers' rationality, and the consequences of those limits for their economic and managerial behavior.

*Simon (1982) argues that complexity is deep in the nature of things, and discovering tolerable approximation procedures and heuristics that permit huge spaces to be searched selectively is at the heart of intelligence, whether human or artificial.* A theory of rationality that does not give an account of problem solving in the face of complexity is sadly incomplete. It is worse than incomplete; such theory can be seriously misleading by providing "solutions" to organizational economic questions that are without operational significance. The theory of heuristic search, cultivated in artificial intelligence and information-processing psychology, is concerned with devising or identifying search procedures that will permit systems of limited computational capacity to make complex decisions and to solve difficult problems. As Franco Modigliani was fond of saying, "If businessmen are not now maximizers, after enough of them have graduated from business school, they will be." So we might even expect that a positive theory of organizational economic behavior and strategic management will have to include as a sub-theory the way in which business schools produce, and diffuse, decision-making techniques. Procedural rationality is the rationality of a person for whom the time and effort required for computation are scarce human resources.

Simon (1982) concludes by noting that there is a saying in politics that "you can't beat something with nothing." You can't defeat a measure, or a candidate, simply by pointing to defects and inadequacies. You must offer an alternative. What then is the status of the neoclassical economic theory of the firm? There can no longer be any doubt that the micro-

analytic assumptions of neoclassical economic theory -- the assumptions of perfect rationality -- are contrary to fact. It is not a question of approximation; the assumptions of perfect rationality do not even remotely describe the processes that humans use for making decisions in complex business situations.

Moreover, there is an alternative. If anything, there is an embarrassing richness of alternatives. Today, we have a large mass of descriptive data from both laboratory and field, which show how human problem solving and decision-making take place in a wide variety of situations. A number of theories incorporate the replacement of optimization by targets and satisficing goals, and mechanisms of learning and adaptation. If our research interest is in descriptive decision theory (or even normative decision theory), it is now clear that the neoclassical economic theory of the firm has been challenged by a superior alternative that provides researchers with a much closer approximation to what is actually going on here.<sup>2</sup>

Now that we have studied the “Behavioral Theory of The Firm” from the Carnegie School, we turn next to Chapter 2 concerning transaction costs theory. Oliver Williamson was a doctoral student at Carnegie in the 1960s and we shall see the influence of Richard Cyert, James

---

<sup>2</sup> For further theory development and applications of the Carnegie School, see Allison (1971), Earl (2001), Gimeno, Folta, Cooper and Woo (1997), Mahoney (1992a), March (1988, 1999), Scott (1987), Simon (1957, 1996), and Thompson (1967). Several consequences of bounded rationality have been observed including: (1) selective perception of information; (2) processing of information is done in an adaptive, sequential manner; (3) heuristic procedures reduce mental effort; and (4) memory works by a process of active reconstruction. Systematic biases result with insensitivity to prior probability of outcomes; insensitivity to sample size; misconceptions of chance; failure to recognize regression toward the mean; biases due to the retrievability of instances; illusory correlation; insufficient adjustment and anchoring; and biases in the evaluation of conjunctive and disjunctive events (Kahneman, Slovic and Tversky, 1982). Given the limitations and (systematic) biases of the individual, those operating from a behavioral perspective tend to view the organization as a more efficient information processor than any given individual. The firm is considered to be an institutional response to uncertainty and bounded rationality --- a theme that we will see below in transaction costs theory.

March, and especially Herbert Simon on Williamson's (1975, 1985, 1996) transaction costs theory. Indeed, transaction costs theory combines their works with economics and aspects of the law in an effort to re-conceptualize the problem of economic organization. Organization theory supports transaction costs theory in terms of: (1) insisting that workably realistic behavioral assumptions are an alternative to the assumptions of economics, which are typically chosen for analytical convenience; (2) the autonomous adaptation of the market is joined with cooperative adaptation by organizations; and (3) the embeddedness (e.g., the institutional environment) of a transaction matters (Granovetter, 1985; North, 1990). Before analyzing Williamson (1975), however, we begin the next chapter with Arrow's (1974) *The Limits of Organization*, which is followed by Coase (1988).

## Chapter 2                      Transaction Costs Theory

The origin of transaction costs theory is Coase's (1937) classic journal article on the nature of the firm. However, it took until the mid-1970s for transaction costs theory to become influential in both research and public policy following the works of Arrow (1974) and especially Williamson (1971, 1979). This chapter covers Arrow (1974), Coase (1988), and Williamson's three transaction costs books (1975, 1985, 1996).

Arrow's (1974) book on *The Limits of Organization* was originally given as the Fels Lecture for 1970-71 to the Fels Center of Government at the University of Pennsylvania. This lucidly written book focuses on at least four major themes: (1) the concept of rationality (individual and social); (2) information economics; (3) the agenda of organizations; and (4) the concepts of authority and responsibility. Throughout this work, Arrow (1974) considers the (often conflicting) demands of society and the needs of the individual and insists that some sense of balance is required.

Coase's (1988) book on *The Firm, the Market and the Law* is a collection of his journal articles and economic writings. Coase (1988) argues that if we move from a regime of zero transaction costs to one of positive transaction costs, what becomes immediately clear is the crucial importance of the legal system. Coase (1988) maintains that it makes little sense for economists to discuss the process of exchange without specifying the institutional setting within which the trading takes place, since the institutional setting affects the economic incentives to produce and the economics costs of transacting.

Williamson's (1975) *Markets and Hierarchies* is a work of scholarship written for posterity. This research book combines ideas from Commons (1934), Coase (1937), Barnard

(1938), Simon (1947), March and Simon (1958), Chandler (1962), and Arrow (1974), among others. In particular, Williamson (1975) emphasizes the fundamental importance of information (Arrow, 1974), organizational innovation (Chandler, 1962), transaction costs (Coase, 1937), and behavioral assumptions (Simon, 1947). Simultaneously, Williamson (1975) provides original ideas concerning vertical integration and the theory of the firm that have proven to be fruitful for the evolving science of organization.

Students of strategic management are well advised to study closely transaction costs theory. Transaction costs theory combines logical rigor with practical relevance to help describe, explain, and predict governance based on comparative efficiency criteria. In addition, the habits required of good scholarship can be learned from Williamson (1975):

- Describe others' works fairly and accurately;
- Search extensively to find and develop plausible theory to help understand the phenomena at hand;
- Have an active mind;
- Work from a disciplinary base; and
- Have the courage to be interdisciplinary, if so inclined.

The organizational economics principles found in transaction costs theory are durable. Indeed, we are currently witnessing greater vertical de-integration (e.g., strategic outsourcing), arguably, as a result of fundamental transaction costs changes --- impacting input and output measurement costs and asset specificity (Mahoney, 1992c) --- that are due to dramatic changes in the development and diffusion of information technology (Shapiro and Varian, 1999). I believe this research area to be especially fruitful for students in the next generation of strategic management research.

Williamson (1985) documents well that empirical research testing transaction costs theory has been largely corroborative. Nevertheless, like all empirical research studies, students today will continue to be challenged by potential specification problems, measurement problems, and identification problems in econometric testing. Given the current state of the empirical research literature testing transaction costs theory (and even more so in other research areas of organizational economics), students in the next generation can arguably do better.

Williamson (1996) continues the agenda of joining law, economics, strategic management and organization theory. A systems perspective is employed to look at (incomplete) contracting in its entirety. Credible commitments are viewed as ways of safeguarding against contractual hazards inherent in incomplete contracts.

Like Barnard (1938), and Simon (1947), Williamson's (1975, 1985, 1996) transaction costs trilogy provides a conceptual framework simple enough to be used and complex enough to accommodate continuing insights into the workings of organization. Students of strategic management taking an organizational economics approach should take heart in achieving and pursuing the science of organization.

We begin this chapter with Arrow (1974). Arrow (1974) and Williamson (1971, 1975) were among the earliest economists to build upon Coase (1937, 1960).



Arrow, Kenneth J. (1974). The Limits of Organization. New York, NY: W. W. Norton and Company.

Rationality: Individual and Social. Arrow (1974) states that the intricacies and paradoxes in relations between individuals and their actions in social contexts have been put well by the sage, Rabbi Hillel: "If I am not for myself, then who is for me? And if I am not for others, then who am I? And if not now, when?" *Here we have, in three successive sentences, the essence of a tension that most feel between the claims of individual self-fulfillment, and the claims of social conscience and action.* Tensions between society and the individual are inevitable. Their claims compete within the individual conscience, as well as in the area of social conflict.

Arrow (1974) discusses the relationship between society and the individual in the rational spirit of the economist. Arrow (1974) notes that economists by training think of themselves as guardians of rationality, ascribers of rationality to others, and the prescribers of rationality to the social world. It is this social science role that Arrow (1974) chooses to play.

Arrow (1974) observes that a truly rational discussion of collective action in general or in specific contexts is necessarily complex, and such a discussion is necessarily incomplete and unresolved. Arrow (1974) takes the (logical positivist) position of Simon (1947), and maintains that rationality has to do with means and ends and their relation. The concept of rationality does not specify what the ends are.

The role of an economist is sometimes unpleasant. It's probably not entirely accidental, though a little unfair, that Carlyle referred to economists as the practitioners of the dismal science. Economists frequently have to point out the limits of our opportunities. Economists have to say, "This or that, not both. You can't do both." Moreover, economists have to point out

frequently that the economic system is complex in its nature. It can easily happen that a step that on the face of it is an “obvious” way of achieving certain desired values may, in fact, frequently lead to their opposites. Arrow (1974) cites, for example, many proposals for high increases in minimum wages. Surely, most would want to redistribute income to the lower end of the wage scale. The most obvious thing to do is to raise the wages. Economists realize that the situation is not that simple, that the system can react to that policy; the system does not passively accept such a change. The end result may be a substitution of capital for labor and consequently an increase in the volume of unemployment, an economic outcome that is arguably worse than low wages.

The basic resources of society -- its natural resources, its human resources, and its technological resources -- are limited in supply, and the realization of alternative values or the search to find alternative activities for meeting those values imply a competition for these scarce resources. If we do things one way, we cannot do them another way. So we need to have a system that mediates this competition for the services of scarce resources, whether it is a market or an authoritative allocation system, as in the military or in the socialist state.

Further, interpersonal organization is needed to secure the economic gains that can accrue from cooperation. The essential considerations are two: (1) individuals are different and in particular have different talents, and (2) individuals' efficiency in the performance of social tasks usually improves with specialization. We need cooperation in order to achieve specialization of function. This cooperation and specialization involves all the elements of trade and the division of labor (Stigler, 1968).

How do we evaluate alternative social organizations? Again, for a commonplace of economic thought, we use the concept of efficiency or optimality that is associated with the name of Vilfredo Pareto. Whatever else we mean by better or efficient, we certainly mean the following: one situation, one system, or one allocation is better than another if every individual feels it is better according to his or her own individual values. Under specific assumptions it is shown that efficiency can be achieved through a particular kind of social system, the price system. Although not mentioned by Arrow (1974), a state in which some people are starving and suffering from acute deprivation while others are tasting the good life can still be Pareto optimal if the poor cannot be made better off without cutting into the pleasures of the rich -- no matter by how small an amount. Pareto optimality is faint praise indeed.

Arrow (1974) notes that in a price system, individuals have a sense of freedom. Individuals are free to act within the system; there is no direct order telling individuals what to do. Individuals have an income and they can spend it. Needless to say, this freedom of action is, from a certain point of view, somewhat illusory. This freedom of action can be very small indeed if their income is very low. One's income is not determined by justice, but by a complicated system of interactions where the ethical meaning is difficult to define. *The idealization of freedom through the market completely ignores the fact that this freedom can be, to a large number of people, very limited in scope.*

The price system can also be challenged on the grounds that the price system harnesses self-serving motives that our ethical systems frequently condemn. The price system makes a virtue of selfishness. We should, on the other hand, not ignore the enormous economic gains in efficiency that can be achieved through the price system, as compared with most conceivable alternatives.

Nevertheless, there are profound difficulties with the price system, even within its own logic, and these difficulties strengthen the view that, *valuable though the price system is in certain realms, the price system cannot be made the complete arbiter of social life. The price system does not, in any way, prescribe a just distribution of income.*

There are other drawbacks, which can be discussed in a somewhat more precise way. In fact, in a strictly technical sense, the price system does not always work. You simply cannot price certain things. A classic example of considerable importance is the pollution of water and air.

Arrow (1974) observes that trust has an important pragmatic value. Trust is an important lubricant of a social system. Trust is extremely efficient; it saves a lot of trouble to have a fair degree of reliance on other people's word. Unfortunately, trust is not a commodity that can be purchased very easily. If you have to buy trust, you already have some doubts about what you have purchased. Trust and similar values, loyalty, or truth telling, are examples of what economists call (positive) "externalities." These positive externalities have real, practical, economic value; they increase the efficiency of the system, and they enable people to produce more goods or more of whatever values are held in high esteem. But trust and similar values are not commodities for which trade on the open market is technologically possible or even meaningful.

From the perspective of efficiency, as well as from the perspective of distributive justice, something more than the market is called for. Other modes of governing the allocation of resources are needed. Most conspicuous among these governance modes is the government at all its levels. Government influences the allocation of resources by means that operate within the

price system, but also otherwise. Government has its host of laws and regulations, coercive and certainly non-market methods of controlling and directing the economy and indeed society in general. Government's role in internalizing externalities is then straightforward in principle, which does not imply that it is easy in practice.

With regard to distributive justice, the root facts here are the incommensurability and incomplete communicability of human wants and human values. George Bernard Shaw long ago observed, "Do not do unto others as you would have them do unto you. They may have different tastes."

Conflict, to be sure, is mitigated by the essential human feeling of sympathy, a sense of feeling oneself to be in the other person's place. This motive operates with some, though doubtless inadequate, strength and this motive can operate better in an institution such as the government, designed to give some scope to expressing altruistic interests.

A firm, especially a large corporation, provides another major area within which price relations are held in partial abeyance. The internal organization is hierarchical and bureaucratic. Prices no doubt have powerful influences from the outside, and in many firms there are concerted attempts to simulate the operations of a market, perhaps even to do better than the sluggish and imperfectly informed markets of economic reality. But internally, and especially at lower levels, the relationships among the employees of a firm are different from arm's length bargaining. As Simon (1947) has observed, an employment contract is different in many ways from a commodity contract; an employee is willing to obey *authority*. The employee is free to leave, but since the transaction costs of leaving are always present and frequently non-trivial, the employment relationship creates an expectation of continued participation.

There is still another set of “institutions” that Arrow (1974) calls to our attention. These are invisible institutions: the principles of ethics and morality. Societies in their evolution have developed implicit agreements to certain kinds of regard for others, agreements that are essential to the survival of the society or at least contribute greatly to the efficiency of its working. It has been observed, for example, that among the properties of many societies whose economic development is backward is a lack of mutual trust. Collective undertakings of any kind, not merely governmental, become difficult or impossible. And it is clear that this lack of social consciousness is a distinct economic loss in a very concrete sense, as well as a loss in the possible well running of a political system.

The fact that we cannot mediate all our responsibilities through prices, through paying for them, makes it essential in the running of society that we have what might be called "conscience," a feeling of responsibility for the effects of one's actions on others. Unfortunately, this philosophical position cannot be pushed too far. We cannot know all the effects of our actions on all other people. When you take these obligations to others seriously you are forced into a difficult (existential) position where you take actions whose consequences you cannot really know and yet you feel responsible for these actions.

Social demands may be expressed through formal rules and authorities, or social demands may be expressed through internalized demands of conscience. Looked at collectively, these demands may be compromises that are needed to increase the efficacy of all. At any moment these demands are apt to be felt by the individual as a set of shackles. And, unfortunately, there are still further problems. A commitment to a war or to a revolution or to a religion is typically one that is hard to reverse, even if conditions have changed from the time

when the commitment started. Even if experience has shown the unexpectedly undesirable consequences of a commitment, the past may continue to rule the present. It is such thinking, Arrow (1974) argues, that gives rise to the greatest tragedies of history, this sense of commitment to a past purpose that reinforces the original agreement precisely at a time when experience has shown that such a commitment must be reversed.

Rationality and foresight are indeed capable of creating delay and doubt; so, too, are conscience, respect for others, and the sense of vague respect to distant and unanticipated consequences that we may worry about. There are no simple answers here and Arrow (1974) does not give any. There are moments of history when we simply must act, fully knowing our ignorance of possible consequences, but to retain our full rationality we must sustain the burden of action without certitude, and we must always keep open the possibility of recognizing past errors and changing course.

Organization and Information. Arrow (1974) submits that organizations are a means of achieving the benefits of collective action in situations in which the price system fails. The term "organization," is interpreted quite broadly. Formal organizations, firms, labor unions, universities, or government, are not the only kind. Ethical codes and the market system itself are to be interpreted as organizations. The purpose of organizations is to utilize the fact that many (virtually all) decisions require the participation of many individuals for their effectiveness.

There is one particular failure of the price system to which Arrow (1974) stresses, one that is central to the understanding of organizations. Arrow (1974) refers to the presence of uncertainty. Uncertainty may be about conditions of production or tastes that, if known, would affect individuals' desires to trade. Then, instead of contracts to buy and sell fixed amounts of goods, it would be better to have conditional contracts, or contracts in contingent commodities.

But the range of contingencies for which conditional contracts are available is much more limited than would be ideally desirable in theory. The taking of desirable economic risks is inhibited by the inability to insure against business failure, for example.

There is more than one reason for the failure of the theoretically desirable contingent prices to exist. One reason is the sheer complexity of the price schedule. An insurance policy would have to specify an enormous number of contingencies with, in general, different payments for each possibility. Drawing up such contracts would be expensive. Moreover, the courts of law, on the basis of long experience, have shown little faith in the ability of the average individual to understand complicated contracts.

Another major reason for limitation of the price system for allocating risk-bearing is the difficulty of distinguishing between genuine risks and failures to optimize, a difficulty known as *moral hazard*. For example, the outbreak of a fire may be due to a combination of exogenous circumstances and individual choice, such as carelessness or, in the extreme case, arson. Hence, a fire insurance policy creates an economic incentive for individuals to change their behavior, and ceases to be a pure insurance against an uncontrollable event.

Consider also the problem known in the insurance literature as *adverse selection* (Arrow, 1971). The insured may know their risks better than the insurer, for example, in life insurance. The insurer may start by choosing rates on some actuarial basis. But then the high-risk groups will buy more of the insurance than the average, while the low-risk groups will buy less. Hence, the experience of the insurer, as weighted by dollars, will be less favorable than the actuarial. The rates will have to be raised, but this increase in rates drives still more of the low-risk groups out. A situation will be created in which there are many whose risks are inadequately covered,



because it is not known how low those risks really are. The essential cause of market imperfections in this case is asymmetric information between the two contractual parties. Akerlof (1970) provides a mathematical model where adverse selection can lead to such market failure in the "market for lemons" for used cars.

Another illustration of asymmetric information among economic agents is the relation between patient and physician. It is of the essence of this relationship or other relations between principal and agent that they differ in their information about the world. But this information asymmetry means that there can really be no contract that insures against the agent's failure to do business properly. Arrow (1974) argues in the context of medical economics that one might regard professional ethics as an example of an institution that fills, in some measure, the gap created by the corresponding failure of the price system.

Consider the organization as a processor of information. The scarcity of information-handling capability is an essential feature for the understanding of both individual and organizational behavior. The transformation of probabilities due to signals is precisely what constitutes the acquisition of information.

This definition of information is qualitative, and so it will remain for Arrow's (1974) purposes. The quantitative definition that appears in information theory is probably of only limited value for economic analysis, for reasons pointed out by Marschak (1968); different bits of information, equal from the viewpoint of information theory, will usually have different economic benefits or costs. Thus, let statements A and B be any two statements about the world, for neither of which is its truth or falsity known *a priori*. Then a signal that A is true conveys exactly as much "information" as a statement that B is true. But the economic value of knowing whether or not A is true may be vastly greater than the economic value of knowing B's truth-

value, or it may be that the resources needed to ascertain the truth-value of A are much greater than those for B. In either case, the information-theoretic equivalence of the two possible signals conceals their vast economic difference. There is little that one can say systematically about the economic benefits for information in general. The main remark that can be ventured on is that there are increasing returns to the *uses* of information.

Let us now turn to the economic costs of information. First and most important, individuals themselves are inputs. Immediately or ultimately, the information must enter their brains through their sensory organs, and both brains and senses are limited in capacity. Information may be accumulated in files, but this information must be retrieved to be of use in decision-making. The psychological literature has many empirical research studies of the limits on the sensory perception abilities of humans, and some on their limits as information processors. Individuals' limited capacity for acquiring and using information is a fixed factor in information processing, and one may expect a sort of diminishing returns to increases in other information resources. Organization theorists have long recognized limits of this kind under the heading of the "span of control."

A second key characteristic of information costs is that information costs are in part capital costs, and more specifically they typically represent an irreversible investment. Arrow (1974) does not place much weight on the physical aspects of communication, telephone lines and the like. Rather, Arrow (1974) emphasizes the need for having made an adequate investment of time and effort to be able to distinguish one signal from another. Learning a foreign language is an obvious example.

Now by its very nature the economic value of an (irreversible investment) information-channel is uncertain, and so we have an economic problem that resembles the demand for inventories under conditions of uncertainty. We may venture on some possible generalizations. One is that the demand for investment in information is less than the demand would be if the economic value of the information were more certain. The second generalization is that the random accidents of history ("initial conditions") will play a bigger role in the final equilibrium. Once the investment has been made and an information channel acquired, this information channel will be cheaper to keep on using this information channel than to invest in new channels. Thus, it will be difficult to reverse an initial commitment in the direction in which information is gathered.

A third basic characteristic of information costs is that information costs are not uniform in different directions. At any given moment, individuals are bundles of abilities and accumulated information. Individuals may easily find it cheaper to open certain information channels rather than others in ways connected with these abilities and this knowledge (exhibiting path dependencies). It is also easier to communicate with other individuals with whom one has a common approach or a common language.

The Agenda of Organizations. Arrow (1974) notes that in neoclassical microeconomic (maximizing) theory it is implicit that the values of all relevant variables are at all moments under consideration. All variables are therefore *agenda* of the organization, that is, their values have always to be chosen. On the other hand, it is a commonplace of everyday observation and of research studies of organization that the difficulty of arranging that a potential decision variable be recognized as such may be greater than that of choosing a value for it.

Arrow (1974) maintains that the combination of uncertainty, indivisibility, and (idiosyncratic) capital intensity associated with information channels and their use imply: (a) that the actual structure and behavior of an organization may depend heavily upon history, and (b) the very pursuit of efficiency may lead to core rigidity and unresponsiveness to further change.

Decisions are necessarily a function of information. Hence, if it is decided to collect no information relevant to a certain class of decisions, those decisions are non-agenda. A decision area may be *active*, *monitored*, or *passive*. An active area is one in which experiments are performed, signals received from them, and terminal acts chosen as a function of the signals. A monitored area is one in which some experiments are being performed; the signals received convey too little information to take terminal acts, but if appropriate signals are received, it is optimal to make further experiments, which in turn will yield enough information to bring the terminal acts onto the agenda. Finally, a passive area is one in which no experiments are being conducted.

Experience may place an item on the agenda. In William James's term a "coercive fact" may be more persuasive than any speculation about the potential benefits from change. The sinking of the Titanic led to iceberg patrols. No doubt the changes in payoffs may be changes in perceptions rather than in actuality.

In general, the information received by a member of the organization can be transformed into a much smaller volume for re-transmission without losing value for choice of terminal acts. The theory of sufficient statistics is an example of this reduction of information without loss of value. It is this reduction in re-transmission that explains the utility of an organization for

information handling. Since information is costly, it is better, in general, to reduce the internal transmission still further. The efficiency of a channel can be increased by a suitable choice of code. The teaching and learning of codes by individuals are acts of irreversible investments for them. It is therefore also irreversible capital accumulations for organizations. It follows that organizations, once created, have distinct identities, because the costs of changing the code are those of unanticipated obsolescence.

History matters. The code is determined in accordance with the best expectations at the time of the firm's creation. Since the code is part of the firm's organizational capital, the code of a given firm will be modified only slowly over time. Hence, the codes of firms starting at different times will in general be different even if they are competitive firms. The need for codes mutually understandable within the firm imposes a uniformity requirement on the behavior of the participants. They are specialized in the information capable of being transmitted by the codes, so that they learn more in the direction of their activity and become less efficient in acquiring and transmitting information not easily fitted into the code. Hence, the firm itself serves to mold the behavior of its members.

If we think of education as the primary source of new information, then the youngest and newest members of the organization introduce new information into the organization. Thus, we have the possibility of changes in organizational agenda induced by generational changes. More generally, the prime need in organizational design is increasing capacity to handle a large agenda. To the extent that information and its handling are accumulations of personal capital, what is needed is the "circulation of elites," the turnover of decision-makers. More generally, what is needed is a circulation of information and decision rules. Short-run efficiency within a

narrow framework of alternatives may be less important in the long run than a wide compass of potential activities that have real options value.

Authority and Responsibility. Arrow (1974) notes that among the most widespread characteristics of organizations is the prevalence of the authority relationship. Virtually universally, in organizations of any size, decisions are made by some individuals and carried out by others. The fields in which an authority is valid may be limited; and the recipients of orders at one level may have their own field of authority. But within these limits, the giving and taking of orders, having someone tell someone else what to do is an essential part of the mechanism by which organizations function.

The role of authority does vary among organizations. The military is the extreme case, in which authority is all pervasive and essential. The State also exemplifies authoritative behavior in relation to its citizens, particularly with respect to police and legal control. The State shares with firms a more limited kind of authoritative control over employees. Indeed, as Simon (1947) has emphasized, an employment contract is precisely a contract on the part of the employee to accept authority. An employment contract differs, therefore, from a contract to purchase a commodity; what is bought and sold is not a definite objective thing but rather a personal relationship. Within the scope of the contract, the relation between the employer and employee is no longer a market relation but an authority relationship. Of course, the scope of this authority will usually be limited by the terms of the contract, and more fundamentally, the scope of this authority is limited by the freedom with which an employee can leave the job. But since there is normally transaction costs to the exercise of this freedom, the scope of this authority is not trivial.

When either interests or information differ among the members of the organization, the costs of achieving consensus rise, and hence the value of consensus as a mode of organizational decision-making declines relative to that of authority. Despite the vast research literature in this area including theoretical development of the theory of games, we are far from a good understanding. It is certainly clear that the process of bargaining can itself be a costly one, especially when the successive offers and threats take place not in the play world of re-contract but in the real world of economic ruin and the savage destruction of human lives in war.

The aim in designing institutions for making decisions should be to facilitate the flow of information to the greatest extent possible. This design involves the reduction of the volume of information while preserving as much of its value as possible. To the extent that reduction of volume is accomplished by reduction in the number of information channels, we are led back to the superior efficiency of authority.

Organization is a means of handling social functions when the price system fails. Within the firm, the sanctions which authority can use are basically those of hiring and firing. The State employs the sanction of the criminal law. At one level of analysis, this is a suitable answer. These sanctions do operate, and decisions by authority are obeyed, in part, because of the punishments that might otherwise ensue.

*Ultimately, it seems to Arrow (1974), authority is viable to the extent that it is the focus of convergent expectations.* Individuals obey authority because they expect that others will obey authority. Traffic laws, and in particular signal lights, may be obeyed because it is clearly worthwhile to have a system in which everybody obeys them.

Thus, it is important to make authority visible, so that it serves as a coordinating signal. This need for signaling is why external symbols surround authority. The emphasis on convergent expectations as the source of authority implies its fragility.

The efficiency loss due to informational overload is increased by the tendency to filter information in accordance with one's pre-conceptions. It is easier to understand and accept information congruent with previous beliefs than to overcome cognitive dissonance. Political and especially military history from Pearl Harbor to Vietnam is filled with dismal and disastrous examples. To go to an earlier period, when the Titanic began to broadcast for help, the captain of a nearby ship decided that the message must be a hoax; it was well known that the Titanic was unsinkable. For another important perspective on authority, Fromm (1941) emphasizes man's psychological craving for authority.

It is difficult to imagine an organization in which some element of responsibility does not exist, at least in the long run. In the first place, every real organization is of limited scope. Hence, as Hirschman (1970) stressed, exit from an organization is always possible, though possibly at considerable cost. Ultimately, an authority can be held to account for the exit of its organizational members. Disobedience to orders, organized or unorganized, frequently sets limits to authority, and, like many other sanctions, the fear of such disobedience constitutes an internalization of responsibility.

In my view, Arrow (1974) provides an exemplary contribution to the economics of organization. The book combines microeconomic logic, transaction cost economics, and behavioral economics. Now that we have studied Arrow's (1974) classic, we turn next to Coase



(1988) and The Firm, the Market and the Law. This book contains some of Coase's seminal articles (e.g., Coase, 1937 and Coase, 1960).

I have argued elsewhere (Mahoney, 1992a) that there is an isomorphism between the Coase theorem (1960) that in the absence of transaction costs, liability (property rights) rules do not matter for achieving efficient economic outcomes and that in the absence of transaction costs, organizational form (e.g., vertical contracting versus vertical financial ownership) does not matter for achieving efficient economic outcomes (Coase, 1937). In the one paper, Coase (1960) shows that if transaction costs were zero the law had no purpose in serving economic efficiency and in the other paper, Coase (1937) shows that if transaction costs were zero the firm had no purpose in serving economic efficiency. These two papers then may be regarded as stepping-stones on the way to an economic analysis of studying the law and organizations with positive transaction costs. Such an approach is well underway as we shall see from the works of Williamson (1975, 1985, 1996).

Coase, Ronald H. (1988). The Firm, the Market and the Law. Chicago, IL: University of Chicago Press.

Coase (1988) argues that the firm and the market together make up the institutional structure of the economic system. The concept of transaction costs helps explain why the firm exists, and what activities the firm will undertake. Transaction costs include: (a) search and information costs; (b) bargaining and decision costs; and (c) policing and enforcement costs. Coase (1988) posits that firms emerge to organize what would otherwise be market transactions whenever their costs are less than the costs of carrying out transactions through the market.

The Market. Markets are institutions that exist to facilitate exchange, that is, they exist in order to reduce the costs of carrying out exchange transactions. In an economic theory, which assumes that transaction costs are non-existent, markets have no function to perform. In practice, all exchanges regulate in great detail the activities of those who trade in these markets. For anything approaching perfect competition to exist, an intricate system of rules and regulations would normally be needed. Some economists observing the regulations of the exchanges often assume that they represent an attempt to exercise monopoly power and aim to restrain competition. Coase (1988) submits that they ignore or, at any rate, fail to emphasize an alternative explanation for these regulations: that institutions exist in order to reduce transaction costs and therefore to increase the volume of trade and economic value creation.

The Way Ahead. Coase (1988) maintains that without some knowledge of what would be achieved with alternative institutional arrangements, it is impossible to choose sensibly among them. We therefore need a theoretical system capable of analyzing the effects of changes in these institutional arrangements. To do this analysis, it is not necessary to abandon standard economic

theory, but it does mean incorporating transaction costs into the analysis, since so much that happens in the economic system is designed to reduce transaction costs. Thus, not to include transaction costs impoverishes the theory.

The Nature of the Firm. Coase (1988) argues that it can be assumed that the distinguishing mark of the firm is the supersession of the price mechanism. Coase (1988) points out that while economists treat the price mechanism as a coordinating instrument, they also admit the coordinating function of the “entrepreneur,” and it is surely important to inquire why coordination is the work of the price system in one case, and of the entrepreneur in another case. We have to explain the basis on which, in practice, the organizational governance choice between alternatives is influenced.

*Coase (1988) argues that the main reason why it is profitable to establish a firm would seem to be that there is a cost of using the price mechanism.* The most obvious transaction cost of “organizing” production through the price system is that of discovering what the relevant prices are. This transaction cost may be reduced but this transaction cost will not be eliminated by the emergence of specialists who will sell this information. The costs of negotiating and concluding a separate contract for each exchange transaction, which takes place on a market, must also be taken into account. For a series of contracts is substituted one. The contract is one whereby the employee, for certain remunerations (which may be fixed or fluctuating), agrees to obey the direction of an entrepreneur *within certain limits*. The essence of the contract is that it should only state the limits to the control rights of the entrepreneur. Within these limits, the entrepreneur can therefore direct the employees.

Coase (1988) then asks why isn't all production carried on in one big firm? Coase (1988) states that there are at least two important reasons. First, as a firm gets larger, there may be

decreasing returns to the entrepreneur function, that is, the costs of organizing additional transactions within the firm may rise. Second, as the transactions that are organized increase, the entrepreneur fails to place the factors of production in the use where their economic value is greatest, that is, fails to make the best use of the factors of production. These two reasons correspond to the phrase of “diminishing returns to management.”

The Problem of Social Cost. Coase (1988) is concerned here with another aspect of transaction costs economics. Here the focus is on those actions of business firms that have harmful effects on others. The standard example is that of a factory, the smoke from which has harmful effects on those occupying neighboring properties.

*Coase (1988) insightfully explains that we are dealing with a problem of a reciprocal nature. To avoid the harm to B would be to inflict harm on A. The real question that has to be decided is: Should A be allowed to harm B, or should B be allowed to harm A?* Coase (1988) submits that with costless market transactions, the decision of the courts concerning liability of damage would be without effect on the allocation of resources in terms of economic efficiency.

When dealing with the problem of the rearrangement of legal rights through the market, Coase (1988) argues that a rearrangement would be made through the market whenever this change would lead to an increase in the economic value of production. But this argument assumes costless market transactions. Once the economic costs of carrying out market transactions are taken into account, it is clear that such a rearrangement of property rights will only be undertaken when the increase in the economic value of production consequent upon the rearrangement is greater than the economic costs that would be involved in bringing such a change about. In these conditions, the initial delimitation of property rights does have an effect

on the efficiency with which the economics system operates. One arrangement of rights may bring about a greater economic value of production than any other arrangement of rights. But unless this is the arrangement of rights established by the legal system, the costs of reaching the same result by altering and combining rights through the market may be so great that this optimal arrangement of rights, and the greater economic value of production which such an arrangement of rights would bring, may never be achieved.

In summary, the same approach that with zero transaction costs demonstrates that the allocation of resources remains the same whatever the legal position, also shows that, with positive transaction costs, the law plays a crucial role in determining how resources are used. As we shall see later in Chapter 3 on property rights, Coase's (1988) insights have been seminal contributions to not only transaction costs theory, but also property rights theory. We provide a brief application of the Coase Theorem, and we then turn to Williamson's trilogy:

- *Market and Hierarchies* (1975);
- *The Economic Institutions of Capitalism* (1985); and
- *The Mechanisms of Governance* (1996).

---

### **Application of the Coase Theorem: “The Fable of the Bees”**

Source: Cheung, S. (1973), “The Fable of the Bees: An Economic Investigation,” *Journal of Law and Economics*, 16: 11-34.

A prominently discussed case of externalities is the so-called Fable of the Bees. Beekeepers provide pollination services for the surrounding fruit growers, while the fruit growers, in turn, provide nectar for the bees. Many economists would regard this example to be a classic case of (positive) externalities. If beekeepers and growers do not receive compensation for the benefits that they bestow on the other parties, then the beekeepers and growers will under-invest in these activities (from a social welfare perspective).

The Coase Theorem suggests the possibility that beekeepers and growers can privately negotiate with each other, provided the transaction costs are low, and can thus achieve a contractual solution to this (positive) externality problem. Indeed, Cheung (1973) found that this contractual solution is exactly what was done. Beekeepers and growers often enter into contractual relationships. Fruit growers hire hives of bees to provide pollination of those trees that give little suitable nectar, while the beekeeper pays the fruit growers for the privilege of utilizing their bees on high nectar-producing trees. Given these contractual payments, beekeepers and growers have the economic incentives to consider the spillover effects on the other contractual party when they make their investment decisions. Through this market-based process, beekeepers and fruit growers can reach efficient levels of investment with no assistance from centralized (government) coordination.

---

Williamson, Oliver E. (1975). Markets and Hierarchies: Analysis and Antitrust Implications. New York, NY: Free Press.

Williamson (1975) is concerned with the organization of economic activity within and between markets and hierarchies. Market transactions involve exchanges between autonomous economic parties. Hierarchical transactions are transactions for which some form of subordination prevails. Whereas received microeconomic theory generally regards the organization of economic activity between firms and markets as a datum, the study of markets and hierarchies assesses the comparative efficiency properties of alternative governance modes. This theoretical approach is usefully applied to product markets, labor markets, capital markets, and value-chain analysis (Porter, 1985). An organizational failures framework is proposed and employed in an attempt to assess the comparative efficacy of completing related sets of transactions across a market or within a firm.

Following Commons (1934), Williamson (1975) maintains that the transaction is the ultimate unit of microeconomic analysis. Williamson (1975) draws on an extensive market failure research literature that was in place by 1975. Williamson (1975) approaches organizational boundary issues in an interdisciplinary way where law, property rights theory, microeconomic theory, business history, and organization theory are usefully brought together to achieve a better understanding of the origins, evolution, and functions of various firm and market structures.

Williamson (1975) notes that hierarchy usually implies a superior-subordinate relationship. What is called an "employment relationship" is commonly associated with voluntary subordination. Williamson (1975) points out that the question is not merely whether internal organization can be substituted for the market with beneficial results, but what type of

internal organization is to be employed. This second question poses organizational form issues. *Whereas simple hierarchy and vertical integration can be regarded as substitutions of internal organization for failures in the labor and intermediate product markets, respectively, conglomerate organization -- of the appropriate multi-divisional kind -- constitutes the substitution of internal organization for failures in the capital market.*

Williamson (1975) identifies the following themes: (1) while the relation of technology to organization remains important, this relation is scarcely determinative of organizational form. Transactional considerations, not technology, are typically decisive in determining which mode of organization will obtain in what circumstances and why; and (2) more self-conscious attention to rudimentary human attributes is essential if we are to understand more adequately the problems of markets and hierarchies.

Important antecedents of Williamson (1975) are:

- Commons (1924, 1934) made the transaction the ultimate unit of economic analysis, and made transfers of legal control and the efficacy of contracting a primary research focus.
- Coase (1937) where transaction costs are made the centerpiece of an economic analysis of the efficiency of completing transactions as between firms and markets.
- Hayek (1945, 1948) where the adaptive property of competitive market systems under changing market circumstances is emphasized.

[Hayek (1945) highlights the "marvel" of the economic system where prices serve as sufficient statistics, thereby economizing on bounded rationality. Williamson (1975) argues that: Given bounded rationality, uncertainty and idiosyncratic knowledge, prices often do not qualify as sufficient statistics and that a substitution of internal organization (hierarchy) for market-mediated exchange often occurs on this account.]

- Market Failures Literature including analysis of: (1) incomplete markets due to uncertainty; (2) insurance problems; (3) employment relations; (4) vertical integration; (5) capital markets; (6) increasing returns and sunk costs;



(7) indivisibilities; (8) information asymmetries, (9) public goods; and (10) lack of definition of property rights and externalities with positive transaction costs (see Coase, 1960).

Williamson (1975) notes that the transaction costs approach differs from this earlier economics literature since the transaction costs approach is interdisciplinary, combining economics, law, and organization theory. For example, Williamson (1975) combines contingent claims contracting ideas along with "organization man" (Simon, 1947) and with "strategic behavior" (Schelling, 1960). Second, Williamson (1975) is more concerned with tracing the governance structure ramifications of bounded rationality. Third, Williamson (1975) introduces the concept of opportunism and the ways that opportunistic behavior is influenced by economic organization. Fourth, Williamson (1975) emphasizes that it is not uncertainty or small numbers, individually or together, that occasion market failure but it is rather the joining of these factors with bounded rationality on the one hand and opportunism on the other hand that gives rise to exchange difficulties. Williamson (1975) provides then an organizational failures framework:

- Markets and firms are alternative governance modes for completing a related set of transactions;
- Whether a set of transactions ought to be executed across markets or within a firm depends on the relative efficiency of each governance mode;
- The costs of writing and executing contracts vary with objective properties of the market;
- Although the human and environmental factors that impede exchanges between firms (across a market) manifest themselves somewhat differently within the firm, the same set of factors apply to both. A *symmetrical analysis of trading* thus requires that we acknowledge the transactional limits of internal organization as well as the sources of market failure;
- The markets and hierarchies approach attempts to identify a set of *environmental factors* that together with a set of *human factors* explain the circumstances under which complex contingent claims contracts will be costly to execute, and enforce.

Faced with such contractual difficulties, and considering the risks that simple (or incomplete) contingent claims contracts pose, the firm may decide to bypass the market and resort to hierarchical modes of organization. Transactions that might otherwise be handled in the market are thus performed internally, governed by organizational processes, instead. The environmental factors that lead to prospective market failure are uncertainty and small-numbers exchange relations. Unless joined, however, by a related set of human factors, such environmental conditions need not impede market exchanges. Pairing of uncertainty with bounded rationality and joining small numbers with opportunism are especially important.

- The principle of bounded rationality has been defined by Simon as follows: "The capacity of the human mind for formulating and solving complex problems is very small compared with the size of problems whose solution is required for objectively rational behavior in the real world" (1957: 198). Bounded rationality refers to neuro-physiological limits on one hand and language limits on the other.

[If, in consideration of these contractual limits, it is very costly or impossible to identify future contingencies and specify ex ante, appropriate adaptations, long-term contracts may be replaced by internal organization. Internal organization permits adaptation to uncertainty to be accomplished by organizational processes in a sequential fashion. Rather than attempt to anticipate all possible contingencies from the outset, the future is permitted to unfold. Internal organization, in this way, economizes on the bounded rationality of decision-makers in circumstances in which prices are not "sufficient statistics" and uncertainty is substantial.]

- Opportunism involves self-interest seeking with guile. Opportunistic inclinations pose little risk as long as there are competitive (large-numbers) exchange relations. Many transactions that at the outset involve a large number of qualified bidders are transformed in the process of contract execution, so that a small-numbers supply condition obtains at contract renewal. Recurrent short-term contracting is costly and risky when there are opportunistic behaviors and small-numbers. Opportunism may include: (i) ex ante, "adverse selection" (hidden information), (ii) ex post "moral hazard" (hidden action), and (iii) hold-up problems.
- Bounded rationality and uncertainty pose problems for long-term contracting. Opportunism and small-numbers pose problems for short-term contracting. Thus, internal organization may arise because of its relatively greater efficiency.
- Internal organization allows for adaptive sequential decision-making (to economize on bounded rationality) and opportunism is attenuated because: (a) internal divisions do not have preemptive claims on profit streams; and b) internal incentives and controls are much more refined than market exchanges. Thus, the firm is better able to take a long-term view for investment purposes (and hence is more prepared to put

specialized plant and equipment in place) while simultaneously adjusting to changing market conditions in an adaptive sequential manner.

- Whichever way the assignment of transactions to firm or market is made initially, the governance choice ought not to be regarded as fixed. For example, the degree of uncertainty associated with the transactions in question may diminish, market growth may support large-numbers supply relations; and information asymmetries between the parties often shrink. Also, changes in information-processing technology may occur which alter the degree to which bounded rationality limits apply, with the result that a different assignment of activities between markets and hierarchies than was selected initially becomes appropriate later. Thus, we ought periodically to reassess the efficiency of completing transactions by one governance structure rather than another governance structure.

Williamson (1975) notes that the key idea is a concern with comparative governance assessment. Organizational failure is a symmetrical term meant to apply to markets and hierarchies alike. Key concepts include:

Bounded rationality refers to human behavior that is *intendedly* rational but only *limitedly* so (Simon, 1947). Bounded rationality involves neuro-physiological limits on the one hand, and language limits on the other hand. The physical limits take the form of rate and storage limits on the abilities of individuals to receive, store, retrieve, and process information without error. Limited computational capacity prevents comprehensive contracting required for the standard economic welfare theorems on the existence and optimality of a competitive equilibrium to go through (Arrow, 1974). To the extent that internal organization serves to economize on scarce computational capacity, and does not experience offsetting disabilities, internal organization is favored. The advantages of internal organization are especially significant in relation to adaptive, sequential decision-making.

Williamson (1975) notes that language limits refer to the inability of individuals to articulate their knowledge or feelings by the use of words, numbers or graphics in ways that enable them to be understood by others. Despite their best efforts, individuals may find that

language fails them (possibly because they do not possess the necessary vocabulary or the necessary vocabulary has not been devised). Demonstrations and learning-by-doing may be the only means of achieving understanding when such language difficulties arise. Bounds on rationality are pragmatically interesting only to the extent that the limits of rationality are reached -- which is to say, under conditions of uncertainty and/or complexity.

In the absence of either of these conditions, the appropriate set of contingent actions can be fully specified at the outset. An example is the two-dimensional tic-tac-toe game. The problem of ex ante specification of contingent responses for all moves in this game is, in relation to the computational abilities of most adults, relatively simple.

The corresponding chess problem, by contrast, is impossibly complex; ex ante specification of the full decision tree is infeasible. Chess, in this sense, is a trivial game: if the theory of Chess (i.e., the complete tree of possible games) were fully known there would be nothing interesting left to play. However, if the average length of a game is 40 moves there are 10 to the 120th power possibilities to consider. This example is a decision tree that is unimaginably vast. Given unbounded rationality, contingent claims contracting goes through, whatever the degree of complexity to be dealt with. Similarly, given a sufficiently simple environment, bounded rationality constraints are never reached, and comparative governance choices between firm and market are not posed -- not in any pragmatically interesting way at least.

Williamson (1975) notes that when transactions are conducted under conditions of *uncertainty/complexity* in which it is very costly, perhaps impossible, to describe the complete decision tree, the bounded rationality constraint is binding and an assessment of alternative

governance modes, in efficiency respects, becomes necessary. The distinction between deterministic complexity and environmental uncertainty is pragmatically inessential. What may be referred to as "uncertainty" in chess is uncertainty introduced into a perfectly certain environment by inability -- computational inability -- to ascertain the structure of the environment. But the result of uncertainty, whatever its source, is the same: approximation must replace exactness in reaching a decision. As long as uncertainty or complexity is present in requisite degree, the bounded rationality problem arises, and a pragmatically interesting comparative governance structure choice is often posed.

Williamson (1975) observes that internal organization often has attractive properties that permit the parties to deal with uncertainty/complexity in an adaptive, sequential fashion without incurring the same types of opportunism hazards that market contracting would pose. Such adaptive, sequential decision processes economize greatly on bounded rationality. Rather than specifying the decision tree exhaustively in advance, and deriving the corresponding contingent prices, events are allowed to unfold and attention is restricted to only the actual rather than all possible outcomes. As a Bayesian, one considers the sequential process of successively revising a priori probabilities on the basis of new observations. Thus, you can cross your bridge as you come to it rather than phrase your detailed plan in advance, thereby crossing all possible bridges you might conceivably come to.

Williamson (1975) maintains that a further advantage of internal organization is that, as compared to recurrent market exchange, efficient codes are more apt to evolve and to be employed with confidence by the parties. Such coding also economizes on bounded rationality. Complex events are summarized in an idiosyncratic language. Communication systems become effective when these systems employ languages that carry large amounts of meaning with

relatively fewer symbols. Organizations find such things as blueprints and occupational jargon helpful in increasing the efficiency of their communications. Although, in principle, the parties to recurrent market contracts could devise the same language, thereby realizing the same economies, such exchanges are more subject to contractual risks of opportunism -- hence, are less apt to be developed as fully. An additional advantage of internal organization is that internal organization promotes *convergent expectations* (Malmgren, 1961), serving to attenuate uncertainties that are generated when interdependent parties make independent decisions with respect to changing market circumstances.

Opportunism extends the conventional assumption that economic agents are guided by consideration of self-interest to make allowance for more complex behavior. Opportunism involves self-interest seeking with guile. Opportunistic behavior entails making false or empty threats and promises in the expectation that individual advantage will thereby be realized. Examples include selective or distorted information disclosure, and self-disbelieved promises regarding future conduct. Since opportunistic individuals cannot be distinguished *ex ante* from sincere individuals, relying on promises exposes sales contracts, for example, to opportunism hazards during contract execution and at contract renewal. Internal organization may arise because internal organization permits economies to be realized in initial contracting and/or monitoring respects.

Williamson (1975) cautions, however, that merely to harbor opportunistic inclinations does not imply that markets are flawed on this account. It is furthermore necessary that a *small-numbers* condition prevail. Absent a small-numbers condition, rivalry among a large number of bidders renders opportunistic inclinations ineffectual. Contractual parties who attempt to secure

gains by opportunistic posturing will find, at the contract renewal interval, that such opportunistic behavior is non-viable. When, however, opportunism is joined with a small-numbers condition, the trading situation is transformed, entailing contractual problems of "bilateral monopoly."

Williamson (1975) poses the following transactional dilemma: it is in the economic interest of contractual parties to seek terms most favorable to each of them, which encourages opportunistic representations and haggling. The interests of the *system*, by contrast, are promoted if the parties can be joined to avoid both the bargaining costs and the indirect costs (mainly mal-adaptation costs) that are generated in the process. What is of special interest to transaction costs analysis is that while frequently a large-numbers condition appears at the outset, this appearance may be illusory or may not continue into contract renewal stages. If parity among suppliers is changed by first-mover advantages, so that winners of original bids subsequently enjoy non-trivial cost advantages over non-winners, the sales relationship eventually is effectively of the small-numbers variety. This transformation has relevance not only for examining when separable components will be made internally rather than purchased, but also when the workflows between successive individuals will be exchanged under an employment relationship rather than a sales relationship.

Williamson (1975) argues that internal organization enjoys economic advantages of at least three kinds over market modes of contracting in circumstances where opportunism and small-numbers conditions are joined:

- In relation to autonomous contractors, the parties to an internal exchange are less able to appropriate sub-group gains, at the expense of the overall organization (system), as a result of opportunistic representations. Preemptive claims on profits between separate firms are eliminated. Thus, the economic incentives to behave opportunistically are attenuated. Also, the managements of the trading divisions are

more susceptible to appeals for cooperation. Since the aggressive pursuit of individual interests is to the disadvantage of the system, and as the general office can vary present and prospective compensation (including promotions) to reflect non-cooperation, requests to adopt a cooperative mode are apt to be heeded. Altogether, more nearly joint profit maximizing attitudes and results are to be expected.

- Internal organization can be more effectively audited. The auditing advantage of internal organization is attributable to constitutional and incentive differences that operate in favor of the internal governance mode. External auditors are typically constrained to review written records and documents and in other respects to restrict the scope of their investigation to clearly pertinent matters. Internal auditors, by contrast, have greater freedom of action, both to include less formal evidence and to explore the byways into which their investigations lead. The differential improvement of auditing by merged railroad firms relative to auditing by railroad cartels is illustrative (Chandler, 1977).
- When differences do arise, internal organization has an advantage over market-mediated exchange in dispute settling respects. The firm can settle many disputes by fiat and quasi-judicial functions.

Williamson (1975) notes that contractual problems are often posed by information asymmetries and opportunism (e.g., adverse selection and moral hazard). Williamson (1975) emphasizes that information problems can develop even when contractual parties have identical information if third-party enforcers will have trouble obtaining truthful information. This agency problem is avoided by one of two ways: (i) self-enforcing agreements (e.g., Klein and Leffler, 1981; Telser, 1980), or (ii) courts perfectly enforcing contracts. Relative to the agency research literature, transaction costs theory is more self-conscious about imperfect enforcement of contracts.

In summary, Williamson (1975) argues that the advantages of internal organization in relation to markets are:

1. In circumstances where complex, contingent claims contracts are infeasible and sequential spot markets are hazardous, internal organization facilitates adaptive, sequential decision-making that economizes on bounded rationality;

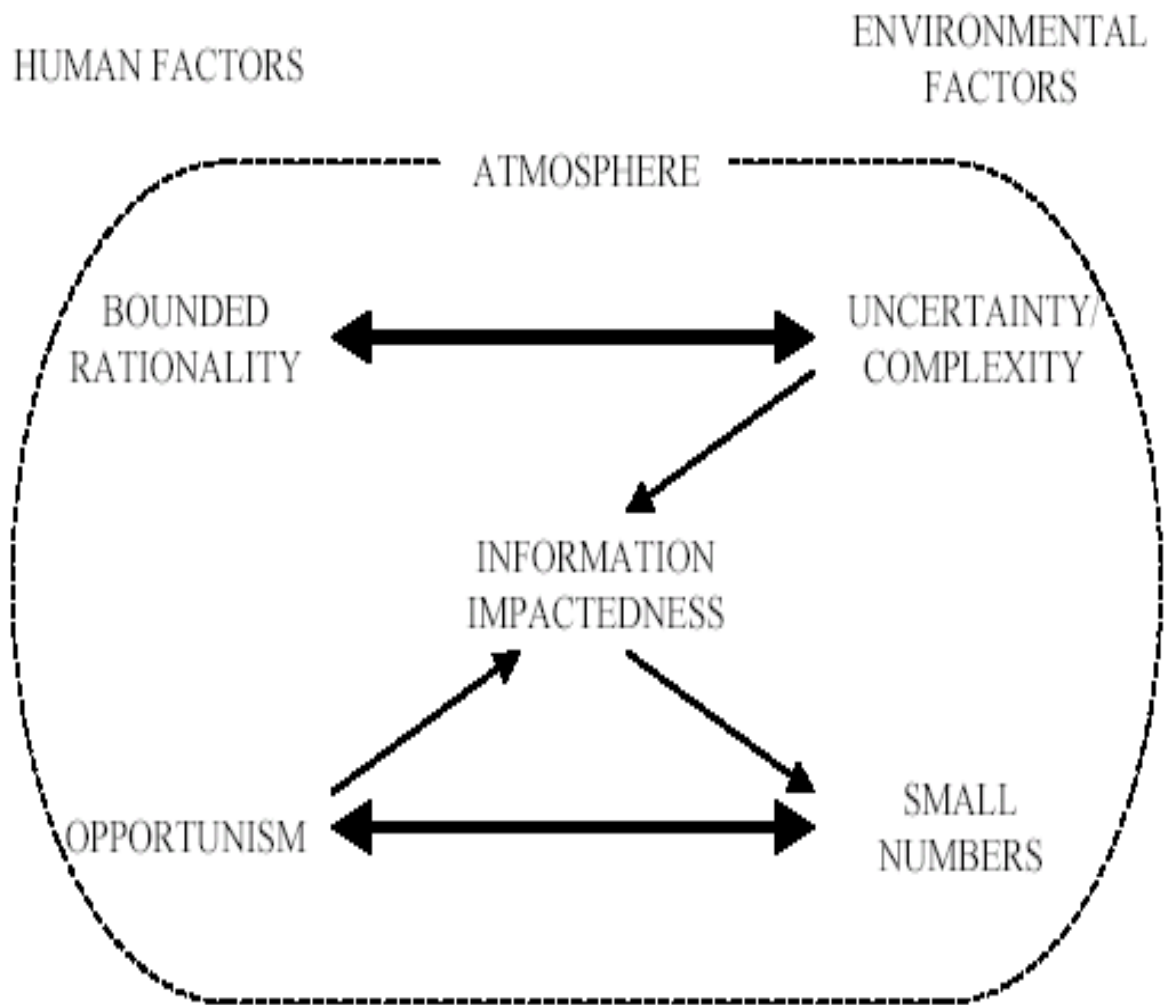


2. Faced with present or prospective small-numbers exchange relations, internal organization attenuates opportunism;
3. Convergent expectations are promoted that reduce uncertainty;
4. Conditions of asymmetric information (information impactedness) are more easily overcome and, when conditions of asymmetric information appear, are less likely to give rise to opportunistic behavior; and
5. A more satisfying trading atmosphere sometimes develops internally.

-----  
Insert Figure 3 from Williamson (1975: 40)  
-----

Simple Hierarchy. Williamson (1975) notes that Alchian and Demsetz (1972) rely on technological non-separabilities as the leading reason to explain the origin of firms. Alchian and Demsetz (1972) use the manual freight-loading example to illustrate worker non-separabilities. Because of interaction effects between workers, the marginal productivity of each worker cannot be determined by observing the total weight loaded during the day. In team production, market mediation is replaced by internal organization where a "boss" monitors the performance of the team and allocates rewards among workers on the basis of observed *input* behavior. Shirking is

**Figure 3. The Organizational Failures Framework**



Source: Williamson (1975: 40).

purportedly attenuated in this way. Williamson (1975) argues that non-separabilities may explain monitoring in some circumstances but non-separabilities hardly qualify as an explanation for complex organizations. Also missing from Alchian and Demsetz (1972) discussion is the

impact of workers' attitudes to their work as the level of monitoring increases (as discussed by the "human relations school of management" in the 1940s, for example).

Understanding the Employment Relation. Williamson (1975) observes that an important form of non-homogeneity --- job idiosyncrasy --- is a key concept for understanding the employment relation. Workers acquire, during the course of their employment, significant firm-specific capabilities, job-specific skills, and related task-specific knowledge.

Williamson (1975) notes that Alchian and Demsetz (1972) take the intellectual position that it is a delusion to characterize the relationships between employers and employees by reference to fiat, authority, or the like. Rather, it is Alchian and Demsetz' (1972: 777) contention that the relationships between employers and employees are identical to that which exists between shoppers and their grocers in fiat and authority respects:

The single consumer can assign his grocer to the task of obtaining whatever the customer can induce the grocer to provide at a price acceptable to both parties. That is precisely all that an employer can do to an employee. To speak of managing, directing, or assigning workers to various tasks is a deceptive way of noting that the employer continually is involved in renegotiation of contracts on terms that must be acceptable to both parties. Long-term contracts between employer and employee are not the essence of the organization we call a firm.

Williamson (1975) points out that implicit in Alchian and Demsetz' (1972) argument is an assumption that the transition costs associated with employee turnover are negligible. Employers, therefore, adapt easily to changing market circumstances by filling jobs on a spot market basis. Although job incumbents may continue to hold jobs for a considerable period of time and may claim to be subject to an authority relationship, all that job incumbents are essentially doing is continuously meeting bids for their jobs in the spot market. Williamson (1975) argues, however, that where tasks are idiosyncratic, in non-trivial degree, the worker-

employer relationship is no longer contractually equivalent to the grocer-customer relationship, and the feasibility of spot-market contracting breaks down. Alchian and Demsetz' (1972) frictionless fiction blurs and neglects essential (incomplete contracting) elements of the organizational economics problem.

Vertical Integration. Williamson (1975) notes that earlier industrial organization research literature on vertical integration emphasized technological interdependency (or in modern terms economies of scope) as an important rationale for vertical financial ownership. The standard example is the integration of iron- and steel-making, where thermal economies are said to be available through vertical integration (Bain, 1968). Williamson (1971, 1975) insightfully (and originally) argues that were it possible to write and enforce a complex contingent claims contract between blast furnace and rolling mill stages, the vertical integration of these activities for thermal economy reasons would be unnecessary. The prohibitive transaction costs of such contracting are what explain the decision to integrate vertically.

Limits of Vertical Integration and Firm Size. Williamson (1975) also analyzes sources of efficiency distortion in internal organization:

- *Internal Procurement Bias.* Sub-goals of a group are easily given greater weight in relation to objective profitability considerations. A norm of reciprocity easily develops;
- *Internal Expansion Bias.* This bias is partly attributable to its dispute-settling characteristics. A common method of dealing with internal system conflict is to adopt a compromise solution where concessions are made to sub-systems rather than require these sub-systems to give up essential functions or resources. This size-preserving tendency is favored by the positive association of both pecuniary and non-pecuniary rewards with size, at least among the functional parts of the organization;
- *Persistence.* If the organizational system has committed itself in advance to the correctness and efficacy of its reform, the organizational system will not tolerate learning of failure.

- *Communication Distortion.* Communication distortion can include (a) serial reproduction loss (a bounded rationality problem); and (b) deliberate distortion (an opportunistic behavior problem).

The Multi-divisional Structure. Williamson (1975) reminds us that Berle and Means (1932: 121) noted that a separation of ownership and control existed and inquired: "have we any justification for assuming that those in control of a modern corporation will also choose to operate it in the interests of the stockholders?" Williamson's (1975) response is that not only have external controls been put in place (e.g., by the market for corporate control) but also internal controls have developed (and as outlined by Chandler (1962) multi-divisional organizations that emerged in the 1920s at General Motors, DuPont, Standard Oil, and Sears, among others, served as internal controls). In particular, the multi-divisional (M-form) structure followed the corporate strategy of diversification (Chandler, 1962). Williamson (1975) emphasizes the following characteristics of the M-form organizational innovation:

- Responsibilities for operating decisions are assigned to (essentially self-contained) operating divisions or quasi-firms;
- The staff attached to the general office performs both advisory and auditing functions. Both functions have the effect of securing greater coordination and control over operating division behavior;
- The general office is mainly concerned with strategic decisions, involving planning, appraisal, and control, including the allocation of resources among the (competing) operating divisions;
- The separation of the general office from operations provides general-office executives with the psychological commitment to be concerned with the overall performance of the organization rather than to become absorbed in the tactical decisions of the functional parts; and
- The resulting M-form structure displays both rationality and synergy: the whole is greater (i.e., more effective and more efficient) than the sum of the parts.

Williamson (1975) argues that in relation to the functional form organization of the same activities, the M-form organization of the large, complex organization served both to economize on bounded rationality and to attenuate opportunism. Operating decisions were no longer forced to the top of the hierarchy but were resolved at the divisional level, which relieved the communication load.

Williamson (1975) notes that imitation of the M-form organizational innovation was at first rather slow. However obvious its superior properties may have been to the organizational innovators, others were naturally skeptical. Organizational innovation may more likely lead to sustainable competitive advantage than product innovation. Organizational innovation may be more complex, not easily articulable and (for a time) tacit.

Williamson (1975) takes issue with the frictionless fiction of perfectly functioning capital markets. Traditional capital market controls are relatively crude because of asymmetric information with respect to internal conditions in the firms and, because of sorting out difficulties, the risk of opportunism on the part of would-be takeover agents is great. Given asymmetric information, outsiders can usually make confident judgments that the firm is not adhering to profit-maximizing standards only at great expense. The large firm is a complex organization, and its economic performance is jointly a function of exogenous economic events, rival behavior, and internal decisions. Causal inferences are correspondingly difficult to make, and hence, opportunism is costly to detect.

Williamson (1975) argues that in a general sense, the most severe limitation of the capital market is that the capital market is an *external* control mechanism. The capital market has

limited constitutional powers to conduct audits, and the capital market has limited access to the firm's incentive and resource allocation fine-tuning mechanisms. The multi-divisional firm can serve as a *miniature capital market*. The multi-divisional firm can:

- *Manipulate Incentives*: Salaries and bonuses can be adjusted to reflect differential operating performance;
- *Perform Internal Audits*: There are at least two advantages of the general office over the capital market in auditing respects. First, division managers are subordinates and both their accounting records and backup files are appropriate subjects for review. Stockholders, by contrast, are more limited in what they can demand in the way of disclosure. Second, the general office can expect knowledgeable parties to be more cooperative than can an outsider, and such cooperation is often rewarded accordingly;
- *Provide Cash Flow Allocation*: The multi-divisional may be thought of as an internal competition among divisions where the general staff can perform a capital market function -- assigning cash flows to high-yield uses. In many respects, this assignment of cash flows to high-yield uses is the most fundamental attribute of the M-form organization in the comparison of internal with external processes. What the multi-divisional lacks in breadth of knowledge the multi-divisional compensates for in its in-depth knowledge of its resources and capabilities. Effective performance of multi-divisional firms with heterogeneous resources is a result not necessarily of having *better* resources but in knowing more *accurately* the relative productive performances of those resources under different business contexts;

Williamson (1975) describes characteristics of effective *divisionalization*:

- Identification of separable economic activities within the firm;
- According quasi-autonomous standing (usually of profit-center nature) to each division;
- Monitoring the efficiency performance of each division;
- Awarding incentives;
  
- Allocating cash flows to high-yield uses; and
- Performing strategic planning (e.g., diversification, acquisition, and related activities).

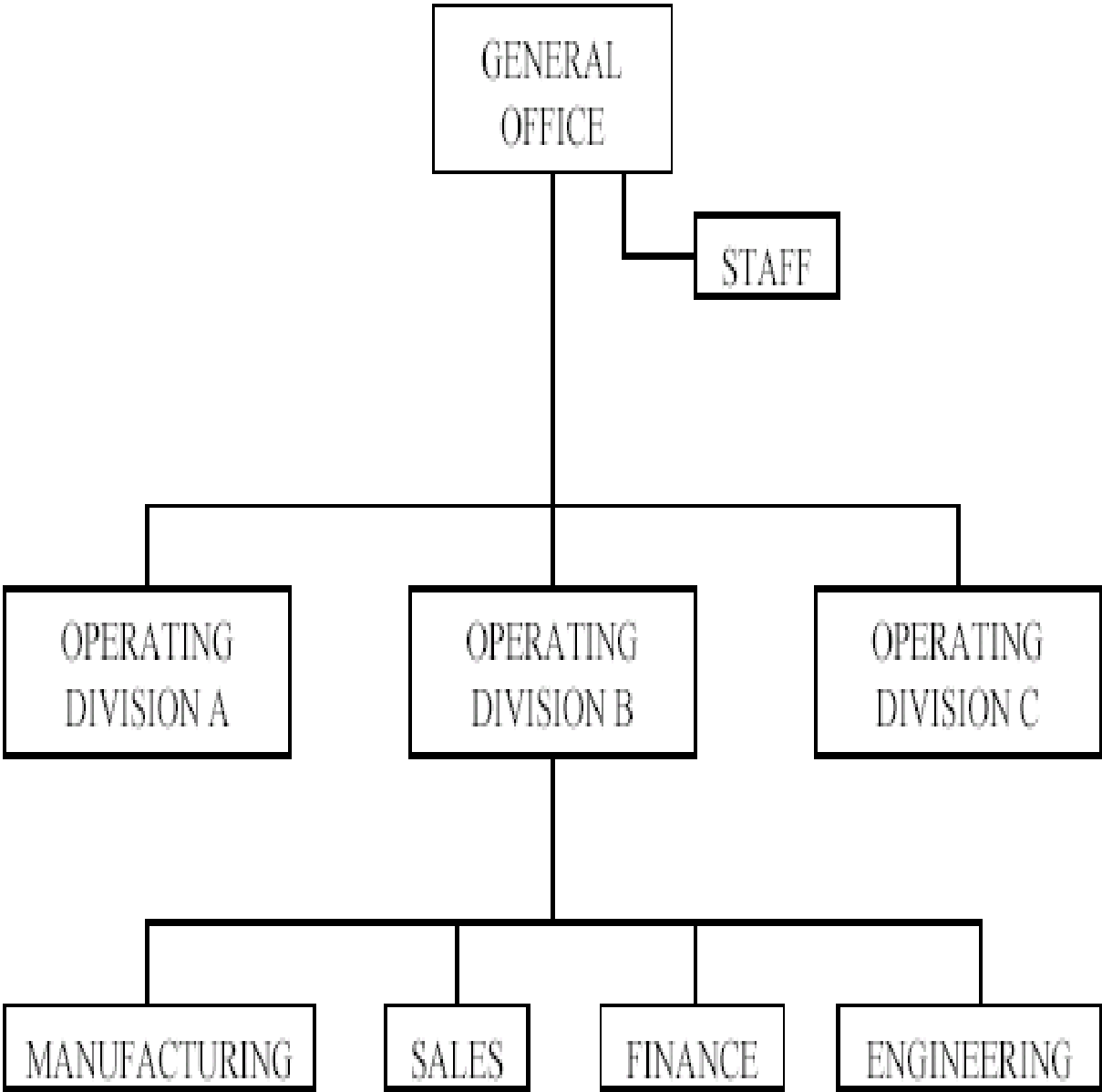
-----  
Insert Figure 4 from Williamson (1975: 138)  
-----

Williamson (1975) emphasizes that care must be exercised lest the general management and its staff become over-involved in operating decisions, and fail to perform the high-level planning and control functions on which the M-form organization relies for its continuing success. Otherwise, “Gresham's law of planning” may occur where daily routine drives out strategic planning.

Williamson (1975) observes that while evolutionary change is to be expected, the hierarchical decomposition principles on which the M-form is based are robust. The concept of hierarchy is given in cybernetics as a necessary structural attribute of any viable organism. All viable systems do in fact exhibit hierarchical organizations. Moreover, not only does Simon's (1962) review of complex biological, physical, and social systems reaffirm this fact, but Simon (1962) emphasizes that hierarchies commonly factor problems in such a way that higher frequency dynamics are associated with the sub-systems, the lower frequency dynamics with the larger systems, and intra-component linkages are generally stronger than inter-component linkages. Hierarchical systems of this sort may be referred to as nearly decomposable. It is not merely fortuitous that the M-form structure factors problems very much in this way. In organization theory these ideas are usually discussed under the heading of "loosely coupled systems.” The importance of sub-systems can hardly be over-emphasized in an analysis of organizational design.

#### **Figure 4. Multidivisional Form**





Source: Williamson (1975: 138)

Williamson (1975) provides the *M-form hypothesis*: The organization and operation of the large firm along the lines of the M-form favors goal pursuit and least-cost behavior more nearly associated with the neoclassical profit maximization hypothesis than does the functional form organizational alternative. The basic argument comes down to this: Just as the vertical integration of production is to be explained in large measure by reference to (comparative governance) failures in the market for intermediate goods, the affirmative aspects of multi-divisional organization are to be understood in terms of (comparative governance) failures in the capital market. Williamson (1975) also points out that pressures due to the market for corporate control induced the more efficient multi-divisional form. Once this multi-divisional form had been extensively established in the 1960s, this organizational structure allowed for greater unrelated diversification as well as a more active takeover market.<sup>3</sup>

Williamson (1975) concludes by highlighting the distinctive features of the *Markets and Hierarchies* approach:

1. The approach makes evident that it is the characteristics of transactions rather than technology that ultimately influences governance choice;
2. A comparative institutional analysis is maintained; markets and hierarchies are regarded as alternative governance modes;
3. The approach makes provisions for rudimentary attributes of human nature (bounded rationality and opportunism) and relates these attributes to a set of environmental factors (complexity/uncertainty and small numbers) in the context of an organizational failures framework;

---

<sup>3</sup> For empirical research in organization theory and strategic management on the multi-divisional form see, for example, Armour and Teece (1978), Mahoney (1992b), Palmer, Friedland, Jennings and Powers (1987), and Rumelt (1974). For example, Mahoney (1992b) finds strong empirical evidence that diversification (as measured by the Berry index, and geographic dispersion (as measured by the log of the number of states/cities in which the enterprise had plants) each induce the adoption of the multi-divisional organizational form.

4. The approach is much more micro-analytic than previous organizational treatments, focusing as it does on the transactional details of recurrent contracting under alternative modes of organization;
5. However useful the fiction of frictionless exchange is for some purposes, it is an impediment to understanding the efficiency properties of alternative modes of economic organization. The frictionless fiction is accordingly abandoned;
6. Organizational form, which is concerned with the decomposition principles of hierarchy, is introduced as an internal organizational counterpart to the familiar market structure measures of industrial organization for explaining economic performance;
7. New questions, or different perspectives on old questions, are afforded across a wide range of strategic management issues -- including the employment relationship, vertical integration, and multi-divisional organization;
8. The approach is comparatively value-free, in terms of the governance choice of markets and hierarchies -- the approach is biased neither for, nor against, unfettered market modes of governance;
9. Supplying a satisfying exchange relation is made part of the economic problem by introducing the concept of "atmosphere." Attitudinal interaction effects are brought to the foreground. Transactions that affect self-esteem and/or perceptions of collective well-being are those for which attitudinal considerations are especially important; and
10. The employment relationship is the leading instance where the influence of metering intensity on work attitudes needs to be assessed with care. More specifically, efforts to divide the employment relation into parts and assess each separately in strictly calculative, instrumental terms can have, for some individuals at least, counter-productive consequences.

Now that we have completed our studies of *Markets and Hierarchies* (1975), we next consider *The Economic Institutions of Capitalism* (1985), where the concept of asset specificity receives greater attention both theoretically and empirically. Indeed, transaction costs theory emphasizes the condition of asset specificity. The existence of durable, firm-specific assets is held to be widespread and the choice of governance structure is held to be economically important.

Williamson, Oliver E. (1985). The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. New York, NY: The Free Press.

Williamson (1985) dedicates his research book to four teachers: Kenneth Arrow, Alfred Chandler, Ronald Coase, and Herbert Simon. From Arrow (1974), Williamson learned the importance of information and not to shoehorn difficult problems into orthodox boxes. Chandler (1962, 1977) taught Williamson that organizational innovation was an important and often neglected phenomenon that had pervasive ramifications for understanding American industry. Coase (1937, 1960) taught Williamson that transaction costs were central to the study of economic organization, and that such studies should be performed in a comparative institutional manner. Simon (1947, 1957) taught Williamson that behavioral assumptions are important, and not to be intimidated by disciplinary boundaries.

Williamson (1985) notes that, like transaction costs research literature, the recent principal-agent research literature is similarly oriented to the study of contract. But there are real differences as well. The principal-agent literature focuses on the ex ante (or economic incentive alignment) side of contract, and assumes that contractual disputes are routinely referred to courts, and assumes that courts dispense justice effectively (indeed, costlessly). In contrast, transaction costs theory maintains that the governance of contractual relations is primarily achieved through private ordering rather than through legal centralism. Although the importance of ex ante incentive alignment is acknowledged, primary attention is focused on the ex post institutions of contract. A compact statement of the economic problem of organization is to devise contract and governance structures that have the purpose and effect of economizing on bounded rationality while simultaneously safeguarding transactions against the economic hazards of opportunism.

Williamson (1985) argues that understanding the economic institutions of capitalism poses deep and enduring challenges to *law*, *economics*, and *organization*. Transaction costs theory is, by design, an interdisciplinary undertaking. Williamson (1985) argues that the transaction costs approach maintains that the institutions of capitalism have the main purpose and effect of economizing on transaction costs.

A transaction occurs when a good or service is transferred across a technologically separable interface. Transaction costs are analogous to *frictions* in mechanical systems. Transaction costs analysis requires a *comparative* institutional analysis among imperfect organizational alternatives. In particular, transaction costs analysis involves the comparative costs of *planning*, *adapting*, and *monitoring* task completion under alternative governance structures. Furthermore, an important task of effective governance is to reduce opportunistic behavior. The motivation and meaning for those studying deeply the economics of organization is that business practice should not be organized to the advantage of the opportunistic against those who are more inclined to keep their promises.

*Williamson (1985) emphasizes the concept of the "Fundamental Transformation" -- a large-numbers condition at the outset (ex ante competition) is transformed into a small-numbers condition during contract execution and at contract renewal (ex post competition) -- has a pervasive importance for the study of the economics of organization.* Rather than characterize the firm as a production function, transaction costs theory maintains that the firm is more usefully regarded as a governance structure. Non-standard contracting -- customer and territorial restrictions, tie-ins, block booking, and related vertical coordination mechanisms have often been presumed to have an anti-competitive purpose and effect. The transaction costs approach discloses that this formulation is simplistic: Many non-standard or unfamiliar contracting

practices serve legitimate transaction cost economizing purposes. Often the parties are engaged in an effort to devise contractual safeguards that promote more efficient exchange. Commercial equivalents of hostages (i.e., mutual sunk cost commitments) arise in this way.

Transaction Cost Economics. Williamson (1985) argues that firms, markets, and “relational contracting” (Macneil, 1980) (e.g., franchising<sup>4</sup>) are important economic institutions. These economic institutions are also the evolutionary product of a fascinating series of organizational innovations. Williamson (1985) advances the proposition that the economic institutions of capitalism have the main purpose and effect of economizing on transaction costs. Compared to other economic approaches to the study of economic organization, transaction costs theory:

- Is more micro-analytic;
- Is more self-conscious about behavioral assumptions;
- Introduces and develops the economic importance of asset specificity;
- Relies more on comparative institutional analysis;
- Regards the business firm as a governance structure rather than a production function; and
- Places greater weight on the ex post institutions of contract, with special emphasis on private ordering (as compared with court ordering).

Williamson (1985) notes that ex ante and ex post transaction costs are usefully distinguished: *Ex ante* transaction costs include: drafting, negotiating, and safeguarding an agreement. Safeguards can take many forms, the most obvious of which is common ownership. *Ex post* transaction costs include: mal-adaptation costs, haggling costs, running of governance structures where disputes are referred, and the economic bonding costs of effecting secure

---

<sup>4</sup> Key theoretical and empirical papers in the economics and strategic management research literature on franchising include: Carney and Gedajlovic (1991), Caves and Murphy (1976), Lafontaine and Shaw (1999), Michael (2000), Norton (1988), Rubin (1978), Shane (1996), and Sorenson and Sorensen (2001).

commitments. Note that ex post transaction costs are related to the "agency costs" of Jensen and Meckling (1976): (a) monitoring expenditures of the principal; (b) bonding expenditures of the agent; and (c) the residual loss. The *ex ante* transaction costs and the *ex post* transaction costs of contracting are interdependent. These transaction costs must be addressed simultaneously rather than sequentially.

In some sense, there are three economic problems posed in Williamson (1985):

- [1] Holding technology (and production costs) constant, minimize transaction costs;
- [2] Minimize the sum of production costs and transaction costs; and
- [3] Consider the problem where asset specificity also enters the revenue function, while taking account of production and transaction costs.

For the most part, we focus here on the second problem of minimizing the sum of production and transaction costs.

Williamson (1985: 24) provides a tree structure for a "*Cognitive Map of Contract*." For the *efficiency branch* of organizational economics there is the economic incentive branch and the transaction costs branch. For the economic incentive branch, there is property rights theory (see chapter 3) and agency theory (see chapter 4). The economic incentive research literature looks mostly at the ex ante side of contracts:

(a) *Property rights* literature emphasizes that ownership matters, where the rights of ownership of an asset take three parts: the right to use an asset, the right to appropriate returns from an asset; and the right to change the form and/or substance of an asset. The economic problem for resource allocation becomes one of getting the property rights correct.

(b) *Agency literature* emphasizes that principals contract in full awareness of the economic hazards that contract execution by agents poses. For example, although the separation of ownership and control attenuates profit incentives, that is anticipated at the time separation occurs and is fully reflected in the price of new shares (Jensen and Meckling, 1976). The structure of the economic problem holds no surprises; all of the relevant contracting action is packed into ex ante incentive alignments.

*Positive agency theory* emphasizes the minimization of monitoring and bonding costs and asserts that natural selection processes are reliably efficacious.

*Principal-agent theory* is a relatively mathematical literature that features ex ante incentive alignments in superlative degree (Holmstrom, 1978). This line of research is akin to the earlier contingent claims contracting research literature but moves beyond it by admitting contracting complications in the form of asymmetric (private) information. Complex economic problems of incentive alignment are posed (which the contingent claims contracting literature had ignored) if full and candid disclosure of asymmetric information cannot be assumed. In other respects, however, the principal-agent research literature and contingent claims contracting are similar. Both research literatures resolve all the relevant contracting issues in a comprehensive ex ante bargain; and both assume that court ordering is efficacious. Economic efficiency drives the argument.

The *transaction costs branch* also maintains the rebuttable presumption that non-standard forms of contracting have economic efficiency purposes. Greater attention is shifted, however, to implementation at the contract execution stage.

*Governance.* The governance branch is the major emphasis of Williamson's (1985) work. In common with the property rights research literature, transaction costs theory agrees that ownership does matter. In common with agency literature, transaction costs theory economizing acknowledges that ex ante incentive alignment matters. But whereas the property rights and principal-agent approaches work within the tradition of legal centralism, transaction costs theory disputes that court ordering is efficacious. Attention is shifted instead to *private ordering*. Governance is especially important when there is bounded rationality coupled with uncertainty, and opportunism coupled with asset specificity.

*Measurement.* First, there can be difficulty of measuring input. Low task programmability reduces the effectiveness of monitoring workers. Second, there can be difficulty of measuring output. For example, Alchian and Demsetz (1972) highlight the high non-separabilities problem of team production.

Williamson (1985) notes that asset specificity is a key contractual dimension. An awareness of asset specificity can be traced at least to Marshall's (1920) discussion of "quasi-rents." Similarly, Marshak (1968) suggested that there exist almost unique, irreplaceable



research workers, teachers, and (firm-specific) personnel, just as there exist unique choice locations for plants and harbors.

Asset specificity refers to durable investments that are undertaken in support of particular transactions. The redeployability of such investments is much lower in best alternative uses or by alternative users should the original transaction be prematurely terminated, and the specific identity of the parties to a transaction matters in these circumstances, which is to say that continuity of the relationship is valued. Thus, contractual and organizational safeguards arise in support of such transactions. Four types of asset specificity are identified in Williamson (1985):

- Site Specificity (e.g., an electric plant and coal mine);
- Physical Asset Specificity (e.g., specialized tools);
- Human Asset Specificity (e.g., firm-specific knowledge); and
- Dedicated Assets (e.g., body dies produced by Fisher Body for GM cars).

Williamson (1985) also classifies uncertainty into three categories:

- Primary (uncertainty about the state of nature);
- Secondary (lack of effective communication); and
- Behavioral (opportunism).

Williamson (1985) argues that, with frequent transactions, the interaction effects between uncertainty and asset specificity are important in an understanding of economic organization, and empirical analysis of transaction cost features is complicated as a result.

-----  
Insert Figure 5 from Williamson (1985: 79)  
-----

**Figure 5. Efficient Governance**

		Investment Characteristics		
		Nonspecific	Mixed	Idiosyncratic
Frequency	Occasional	Market governance (classical contracting)	Trilateral governance (neoclassical contracting)	
	Recurrent		Bilateral governance (relational contracting)	Unified governance

Source: Williamson (1985: 79)

Vertical Integration.<sup>5</sup> Williamson (1985) argues that the main economic differences

between market and internal organization are these:

- Markets promote high-powered economic incentives and restrain bureaucratic distortion more effectively than internal organization;
- Markets can sometimes aggregate demands to realize economies of scale and scope. Therefore, market procurement has advantages in both scale and governance respects where optimal asset specificity is slight; and
- Internal organization has access to distinctive governance mechanisms and enjoys the advantage where asset specificity is substantial.

Williamson (1985) and Williamson and Masten (1999) note that much of the empirical evidence in the research literature that tests transaction costs theory is corroborative, including:

- *Statistical Models Using Field Data:*

[E.g., Anderson, 1985; Anderson and Schmittlein, 1984; Bensaou and Anderson, 1999; Coles and Hesterly, 1998; Heide and Miner, 1992; John and Weitz, 1988; Lyons, 1994; Monteverde and Teece, 1982; Poppo and Zenger, 1998, 2002; Subramani and Venkatraman, 2003; Walker and Poppo, 1991; and Walker and Weber, 1984, 1987];

- *Statistical Models Using Secondary Data:*

[E.g. Balakrishnan and Wernerfelt, 1986; Caves and Bradburd, 1988; Leiblein and Miller, 2003; Leiblein, Reuer and Dalsace, 2002; and Levy, 1985];

---

<sup>5</sup> For mathematical models of vertical integration see Balakrishnan and Wernerfelt (1986), Blair and Kaserman (1983), Riordan and Williamson (1985), and Tirole (1988). For conceptual papers on vertical integration, see Harrigan (1984), Klein, Crawford and Alchian (1978), Mahoney (1992c), Phillips and Mahoney (1985), Teece (1980), and Williamson (1971, 1979). For example, Mahoney (1992c) shows that in the absence of transaction costs, both vertical contracting (e.g., tying contracts, resale price maintenance, exclusive territories, and exclusive dealing contracts) and vertical financial ownership would be equivalent for achieving economic value creation (via lower costs, higher revenues, and/or reduction of risks in ways that cannot be duplicated by shareholders). However, in a business world of positive transaction costs, high task programmability (Eisenhardt, 1985), high non-separability (Alchian and Demsetz, 1972) and high asset specificity (Williamson, 1985) is (in combination) expected to lead to vertical financial ownership because of its comparative governance efficiency advantages. For doctoral students studying the economics of organization, empirical studies that measure all three constructs for explaining and predicting governance structures appear warranted.

- *Focused Case Studies and Studies of Contracts:*  
 [E.g., Argyres, 1996; Crocker and Masten, 1988; Crocker and Reynolds, 1993; Goldberg and Erickson, 1987; Hennart, 1988a; Joskow 1987; Mahoney and Crank, 1995; Masten and Crocker, 1985; Palay, 1984, 1985; Pirrong, 1993; and Williamson, 1985, on CATV];
- *Focused Industry Studies:*  
 [E.g., Afuah, 2001; Dyer, 1996, 1997; Lieberman, 1991; Masten, 1984; Masten, Meehan and Snyder, 1989, 1991; Monteverde, 1995; Nickerson, Hamilton and Wada, 2001; Novak and Eppinger, 2001; Pisano, 1989; Richardson, 1993; and Stuckey, 1983]; and
- *Business History*  
 [E.g., Argyres and Liebeskind, 1999; and Chandler, 1962, 1977]

Business History. Consumer and producer durables requiring information aids, credit, and follow-on service, and producer durables requiring the same were likely candidates for forward integration (e.g., Eastman Kodak, McCormick Harvester, National Cash Register, and Singer Sewing Machine) (Chandler, 1977). While the governance branch emphasizes the role of asset specificity for explaining this pattern, the measurement branch emphasizes (positive and negative) externalities. Externality concerns arise in conjunction with a branded good or service that is subject to quality debasement. Whereas a manufacturer can inspect, and thereby better to control, the quality of components and materials it purchases from earlier stage and lateral suppliers, it is less easy to exercise continuing quality controls over items sold to distributors. If quality enhancement (debasement) efforts of distributors give rise to positive (negative) interaction effects, the benefits (costs) that are incompletely appropriated by (assigned to) the

originators, failure to extend quality controls over distribution results in sub-optimization (e.g., free-rider problems in a franchise system).

The Limits of Firms: Incentive and Bureaucratic Features. Williamson (1985) asks: Why can't a large firm do everything that a collection of small firms can do and more? What is responsible for limitations in firm size? Why not organize everything in one large firm? Williamson (1985) provides two main reasons:

- Economies of scale and scope may be sacrificed if the firm attempts to make for itself what it can procure in the market; and
- Governance costs are increased.

Williamson (1985) points out that the first reason is not a thoroughly comparative explanation. If economies of scale are reached by the outside supplier, then the same economies of scale can be preserved upon merger by instructing the supplier to service the market in the future just as in the past. Thus, the fundamental limitation to firm size must turn on the governance costs disabilities of internal organization where asset specificity is insubstantial.

These governance costs disabilities include:

- "Diminishing returns to management" (Attention as a scarce resource);
- "Control loss" (serial reproduction loss; opportunism);
- More prone to logrolling and subject to politicization;
- Internal procurement biases;
- Bureaucratic delays in large firms; and
- Impairment of economic incentives.

Thus, Williamson (1985) holds that *selective intervention*, whereby organizational integration realizes adaptive gains but experiences no losses, is not feasible.

Credible Commitments: Unilateral and Bilateral Applications: Williamson (1985) notes that credible (sunk-cost) commitments and credible (sunk-cost) threats share this common attribute: Both appear mainly in conjunction with irreversible, specialized investments (i.e., sunk costs). But whereas credible commitments are undertaken in support of joint ventures (Hennart, 1982, 1988b) and strategic alliances (Hennart, 1993; Koja and Prescott, 2002; Robertson and Gatignon, 1998) to promote exchange, credible threats (e.g., excess capacity as a barrier to entry) appear in the context of conflict and rivalry (Sutton, 1992).

Williamson (1985) reminds us that Schelling (1960) also noted that the exchange of human hostages served incentive purposes in an earlier age. Economic hostages, in our business context, involve asset specificity (i.e., sunk costs). Economic hostages are an important component of self-enforcing agreements. *Economic hostages have both ex ante (screening) and ex post (bonding) effects, and Williamson (1985) here focuses on the ex post bonding effects.*

For example, franchisers can better assure quality by requiring franchisee investments in specific assets that upon termination imply a capital loss economic penalty larger than can be obtained by the franchisee by cheating. For example, the franchiser may require franchisees to rent (rather than own) the land upon which their outlet is located. This lease arrangement creates a situation where termination can require the franchisees to move and thereby impose a capital loss on the franchisees up to the amount of their initial non-salvageable investment. Hence, a form of economic collateral to deter franchisee cheating is created. The arrangement is tantamount to the creation of economic hostages to restore integrity to an exchange. The use of economic hostages deters franchisees from exploiting externalities, and is often regarded as an imposed (top down) solution. Franchisees are "powerless;" franchisees accept economic hostage

terms because no other contracting terms are available. Williamson (1985) counsels that such power arguments are often based on ex post reasoning. The use of economic hostages can be, and often is, an efficient *systems solution*, and hence is independent of who originates the proposal. It is the franchise system that benefits from the control of externalities. Indeed, if franchisees recognize that the demand externality exists from the outset, if the franchiser refuses to make provision for the externality in the original contract, and if it is very costly to reform the franchise system once initial contracts are set, franchisees will bid less for the right to a territory than they otherwise would. It should not therefore be concluded that perceptive franchisers, who recognize the demand externality in advance and make provision for it, are imposing objectionable ex ante economic terms on unwilling franchisees. Franchisers are taking steps to realize the full economic value of the franchise system. Here, as elsewhere, contracts must be examined in their entirety.

Critical Commentary on Agency Theory. Williamson (1985) points out that enthusiasts of laissez-faire capitalism are loathe to confront, and are sometimes schizophrenic on the subject of, managerial discretion. Focusing on any given time, these enthusiasts commonly deny the existence of managerial discretion. Comparing current practices with the past, however, those same enthusiasts point with pride to the development of new governance mechanisms that have brought managerial discretion under more effective control. Williamson (1985) submits that the problem of managerial discretion, due to the separation of ownership and control, is alive and well. Or put differently, agency problems are pervasive. The continuing tension between management and stockholder interests is evident in the numerous efforts that incumbent managements have taken to protect target firms against takeover.

Williamson (1985) concludes that transaction costs theory assumes risk neutrality and investigates organizational form as a means for minimizing transaction costs. Williamson (1985) notes that organizational forms (e.g., franchising) may also have risk-sharing purposes as well.

Williamson (1985) maintains that formal mathematical models of verbal arguments that lose in the translation are scarcely to be counted as gains. Formalization is not wanted at any cost. Sometimes, however, effort at formalization discloses gaps or ambiguities that the verbal arguments did not. Anomalies and contradictions can, and should, push those who employ transaction costs analysis to develop better models.

Williamson (1985) argues that even holding technology constant, at least three things happen when a transaction is transferred out of the market and is placed under unified ownership:

- Ownership changes (property rights change);
- Incentives change; and
- Governance structures (ability to monitor and reward) change.

*Williamson (1985: 408) maintains that: " `Flawed' modes of economic organization for which no superior feasible mode can be described are, until something better comes along, winners nonetheless."*

We conclude this chapter with an application of transaction costs theory to vertical integration in the aerospace industry (Masten, 1984), followed by *The Mechanisms of Governance*. Here, Williamson (1996) further contributes to the evolving science of organization. Indeed, we learn that one answer to why we observe so many kinds of organization is that contractual hazards come in many forms, for which nuanced governance structures are devised and chosen and/or selected.



---

### **Application: Vertical Integration in the Aerospace Industry**

Source: Masten, S. (1984), "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics* 27: 403-417.

Masten (1984) studied the make-versus-buy decisions for nearly 2,000 components of a major aerospace contractor. The firm made many products for the United States government. The company had to choose between making each component or by subcontracting the component for production by another firm. Transaction costs theory suggests that internal production is more likely when the assets are specific and the uncertainties in contracting are large.

Masten (1984) used two measures of asset specificity for each component. The first measured design (physical asset) specificity, that is, the extent to which the component was used exclusively by the company or could be easily adapted for use by other aerospace firms or by firms in other industries. A transistor would be an example of a non-specific item; a circuit board designed to individual specifications would be an example of a component with high asset specificity. The second measure of asset specificity was site specificity.

Masten (1984) also measured the complexity of the product design, which was intended to proxy for uncertainties in contracting. Consistent with the theory, Masten (1984) found that products that were highly design-specific and highly complex were more likely to be produced internally. When the product was both design-specific and complex, there was a 92 percent probability of internal production. If the product was design-specific but not complex, the probability of internal production was 31 percent. The probability of internal production was only 2 percent when the product was neither design-specific nor complex. For this particular company, site specificity was unimportant for explaining the organization of production.

Williamson, Oliver E. (1996). The Mechanisms of Governance. New York, NY: Oxford University Press.

Williamson (1996: 3) begins with: “Institutions. What are they? How do institutions differ? To what purpose and effect? Where does the action reside? What are the mechanisms? What are the refutable implications? What are the public-policy ramifications? What do the data support?” Williamson (1996) argues that organizational economics:

- Holds that institutions matter and are susceptible to analysis;
- Is different from but not hostile to orthodox microeconomic theory; and
- Is an interdisciplinary combination of law, economics, and organization.

*Williamson (1996) develops the argument that many puzzles of economic organization turn on an examination of the mechanisms of ex post governance.* Williamson (1996) appeals to law (especially contract law) and organization (which is broadly construed to include organization theory, sociology, and political science), as well as economics. Williamson (1996) contends that the main purpose and effect of non-standardized contractual forms are to economize on transaction costs. The identification, explication, and mitigation of contractual hazards are central to transaction costs analysis. The analytical action resides in the details of transactions and governance. Williamson (1996) proposes a logical structure of organization in which the alignment of transactions with governance structures is the source of refutable implications. Farsighted contracting, credible commitments, and contractual hazard mitigation figure prominently in the analysis (Williamson and Winter, 1991).

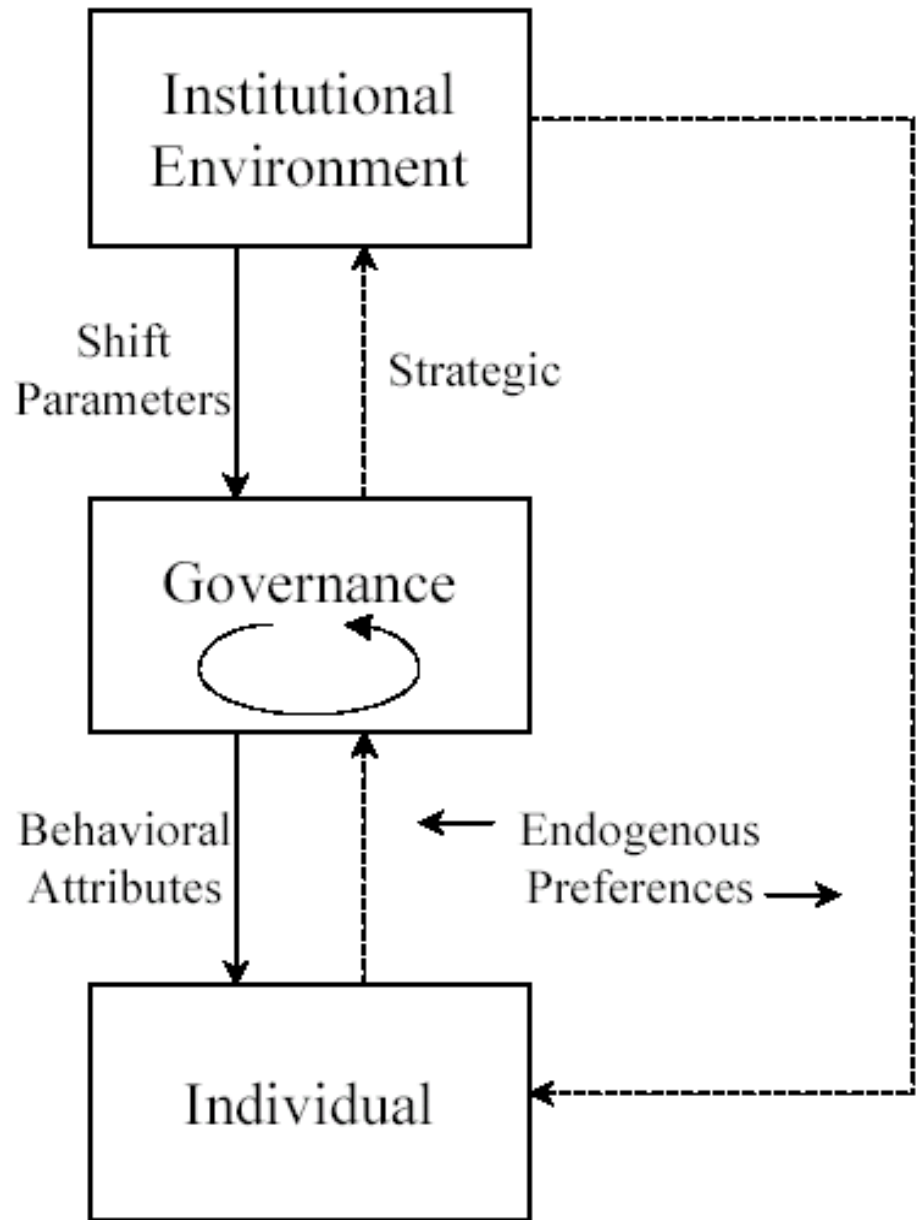
Williamson (1996) submits that the institutional environment (laws, polity, and so forth) and the institutions of governance (markets, hierarchies, and so forth) matter in ways that are pertinent to industrial organization and much else, such as economic history, comparative economic systems, labor economics, business strategy, multinational business, corporate finance,

and organization theory (Scott, 1995). According to North (1990), institutions are the humanly devised constraints that structure political, economic, and social interactions. Institutions consist of both informal constraints (e.g., sanctions, taboos, customs, traditions, and codes of conduct), and formal rules (e.g., constitutions, laws, property rights). North (1990) focuses on the level of the institutional environment, the so-called rules of the game.

Williamson (1996) points out that the second more micro-analytic level at which institutional economics works is at the level of the institutions of governance. Williamson (1996) is primarily concerned with the institutions of governance (e.g., markets, hybrids, hierarchies, bureaus, and so forth). The “institutions of governance framework” mainly takes the institutional environment as given. The institutions of governance operate at the level of individual transactions, whereas the institutional environment is more concerned with composite levels of activity. Another difference is that the two frameworks operate differently with respect to intentionality. Although both the institutional environment and the institutions of governance have evolutionary origins, the ramifications of each are different. The immense difficulties of changing the institutional environment in order to promote economizing outcomes in the aggregate helps explain North's (1990) conclusion that economic history is overwhelmingly a story of economies that failed. By contrast, transaction costs theory contemplates success: Taking the institutional environment as given, economic agents purportedly align transactions with governance structures to effect economizing outcomes. Not only is such a prediction a source of numerous refutable implications, but also the data are largely corroborative.

-----  
Insert Figure 6 from Williamson (1996: 326)  
-----

Figure 6. A Layer Schema



Source: Williamson (1996: 326)

Law, Economics, and Organization. Williamson (1996) submits that parsimony, after all, is what science is after. As with most things, there are trade-offs. Simple theories that finesse or obfuscate core issues are unhelpful. If the action is in the details of transactions and governance, we need to meet the economic problems on terms that are responsive to the needs.

Williamson (1996) notes that orthodox microeconomic theory frequently assumes (often implicitly) that property rights are easy to define, and that the courts knowledgeably enforce property rights and contracts at a negligible cost. In contrast, transaction costs theory treats property rights and contracting as problematic.

Remediableness. Williamson (1996) maintains that transaction costs theory avoids hypothetical ideals and insists that the relevant comparisons are with feasible alternatives, all of which are flawed. Within the feasible options, the relevant economic test is whether an alternative can be described that can be implemented with expected net gains. This test is the remediableness criterion. Claims of (path dependency arguments of) inefficiency (Arthur, 1994), which can be recognized only after the fact and/or cannot be implemented with net gains, have no operational importance.

Williamson (1996) argues that transaction costs theory holds that the main purpose and effect of economic organization is economizing on transaction costs. Transaction costs theory concedes that comprehensive contracting is not a feasible option (by reason of bounded rationality), yet transaction costs theory maintains that many economic agents have the capacities both to learn and to look ahead, perceive contractual hazards, and factor these perceived hazards back into the contractual relation, thereafter to devise responsive institutions. Private ordering through ex post governance is where the main action resides.

Williamson (1996) notes various contractual hazards:

- Hazards of bilateral dependency (Williamson, 1971);
- Hazards that accrue to weak property rights (Libecap, 1989);
- Measurement hazards (Barzel, 1982; Ouchi, 1979); and
- Weaknesses in the institutional environment (North, 1990).

Williamson (1996) submits that superior economic performance is realized by working out of a farsighted but incomplete contracting approach in which the object is to use institutions as (cost-effective) instruments for contractual hazard mitigation.

Williamson (1996) notes that Machiavelli advised his prince to breach contracts with impunity, when by following the contract would be against his interest, and when the reasons that made him bind himself no longer exist. This myopic approach to contract should be contrasted with a more farsighted (but nonetheless incomplete) approach to contract, according to which the prince is advised to mitigate ex post opportunism by crafting ex ante (mutual sunk cost) safeguards. Rather than reply to opportunism in kind, the wise prince is one who seeks both to *give* and to *receive* "credible commitments." Partly that entails *economic* incentive realignment, but mainly the need is to craft governance structures with superior adaptive properties. (See the Appendix to this chapter for a game-theoretic illustration of the importance of crafting ex ante mutual sunk costs safeguards).

Adaptation. Williamson (1996) maintains that adaptation is the central problem of economic organization, of which two types are distinguished: autonomous or Hayekian (1945, 1978) adaptation (in which markets enjoy the advantage) and cooperative or Barnardian (1938, 1948) adaptation (in which the advantage accrues to hierarchy). The study of "incomplete contracting in its entirety" implicates both ex ante incentive alignment and ex post governance.

Williamson (1996) notes that the invisible hand of Adam Smith (1776) and the marvel of the market to which Hayek (1945) referred have spontaneous origins: The price system is one of those institutions that humans have learned to use after they stumbled on it without understanding it. What interests Barnard (1938), however, was not spontaneous cooperation but induced cooperation --- that kind of cooperation that is conscious, deliberate, and purposeful. Barnard (1938) argues that authority is a solution to a complex problem of coordination/adaptation and that such authority arises out of mutual consent. Barnard (1938) maintains that both the decision of an individual to join an organization and the decision to continue reflect a comparative net benefit assessment. Presented with different employment scenarios, persons consciously choose whether or not they will enter into a specific cooperative system. Expanding a zone (of indifference or acceptance) to include greater (potential) burdens or sacrifices must be attended by greater inducements. Barnard (1938) argues that formal and informal organization always and everywhere co-exist, and that informal organization contributes to the viability of formal organization in three significant respects: communication, cohesiveness, and the maintenance of personal integrity. Williamson (1996) notes that left unmentioned by Barnard (1938) was the concept of "influence activities" -- i.e., sub-goal pursuits (see Simon, 1947). Williamson (1996) concludes that to achieve the "science of organization" described by Barnard (1938), economics and organization theory need to come together. Prospects for a science of organization are improved as a consequence.

The Analysis of Discrete Structural Alternatives. Williamson (1996) maintains that: (1) Firms employ different means than markets employ; (2) Discrete contract law differences provide crucial support for, and serve to define each generic form of governance; and

(3) Marginal analysis is typically concerned with second-order refinements to the neglect of first-order economizing.

Williamson (1996) argues that the implicit contract law of internal organization is *forbearance*. Thus, whereas courts routinely grant standing to firms should there be disputes over prices, the damages to be ascribed to delays and failures of quality, courts will refuse to hear disputes between one internal division and another over identical technical issues. Access to the courts being denied, the parties must resolve their differences internally. Accordingly, hierarchy is its own court of ultimate appeal. To review alleged mistakes of judgment or to adjudicate internal disputes would sorely test the competence of courts, and would undermine the efficacy of hierarchy. Accordingly, the reason why the market is unable to replicate the firm with respect to fiat is that market transactions are defined by legal rules of an altogether different kind. There is a logical structure to classical market contracting and there is also a logical structure for forbearance law, and the choice of one regime precludes the other regime. The underlying rationale for forbearance law is twofold: (1) parties to an internal dispute have deep knowledge -- both about the circumstances surrounding a dispute as well as the efficiency properties of alternative solutions -- that can be communicated to the court only at a great cost, and (2) permitting the internal disputes to be appealed to the court would undermine the efficacy of hierarchy. Williamson (1996) notes that this argument contradicts Alchian and Demsetz's claims that the firm has "no power of fiat, no authority, no disciplinary action any different in the slightest degree from ordinary market contracting" (1972: 777). That claim is exactly wrong: firms can, and do, exercise fiat that markets cannot. Prior neglect of contract law differences and their ramifications explain the disparity (see also, Masten, 1988).



Williamson (1996) concludes that neoclassical economics was too preoccupied with issues of allocative efficiency, in which marginal analysis was featured, to the neglect of organizational efficiency ("first order economizing"), in which discrete structural alternatives were brought under scrutiny. Market-favoring predispositions need to be disputed, lest the study of economic organization in all its forms be needlessly and harmfully truncated.

Corporate Finance and Corporate Governance. Williamson (1996) notes that the classical agency theory problem is the separation of ownership and control, and the classical transaction costs problem is vertical integration. Both agency theory and transaction costs theory adopt an efficient-contracting orientation to economic organization. However, there are important differences between agency theory and transaction costs theory:

- Unit of analysis: individual vs. transaction;
- Agency costs focus on ex ante costs; transaction costs emphasize ex post costs;
- There is a legal centralism assumption of agency theory, and a private ordering assumption in transaction costs theory.

Calculativeness, Trust and Economic Organization. Williamson (1996) submits that to craft credible commitments (through the use of economic bonds, economic hostages, information disclosure rules, specialized dispute settlement mechanisms, and the like) is to create functional substitutes for trust. Albeit vitally important to economic organization, such substitutes should not be confused with (real) trust. Transaction costs theory refers to contractual safeguards, or their absence, rather than trust, or its absence. Williamson (1996) argues that it is redundant at best, and can be misleading, to use the term "trust" to describe commercial exchange for which cost-effective economic safeguards have been devised in support of more efficient exchange.

Williamson (1996) suggests reserving the term "trust" for the personal type. A deep and abiding trust relation cannot be created in the face of calculativeness. Calculative trust is a

contradiction in terms. Personal trust is characterized by (1) the absence of monitoring, (2) favorable or forgiving predilections, and (3) discreetness. Such relations are clearly very special. Trust, if it obtains at all, is reserved for very special relations between family, friends, and lovers. Such trust is also the stuff of which tragedy is made. It goes to the essence of the human condition. Not only is "calculated trust" a contradiction in terms, but also user-friendly terms, of which "trust" is one, have an additional cost. The business world is reorganized in favor of the cynics, as against the innocents, when social scientists employ user-friendly language that is not descriptively accurate -- since only the innocents are taken in.

Strategizing, Economizing and Economic Organization. Williamson (1996) notes that the beguiling language of strategizing -- warfare, credible threats, and the like -- notwithstanding, students of economic organization are better advised to focus on issues of economizing (e.g., Barney and Ouchi, 1986; Mahoney, 1992c; Seth and Thomas, 1994; Yao, 1988; and Williamson, 1991) -- of which credible commitments and adaptation are examples. Here, as elsewhere, the need is to get the priorities straight. Williamson (1996) notes that the leading efficiency approaches to business strategy are the resource-based and the dynamic capabilities approach, which he cautiously deems as promising. Williamson (1996) concludes by stating that he anticipates that the science of organization to which Barnard (1938) referred will be realized by this generation of organizational economics scholars.

Now that we have completed our studies on transaction costs theory, where the primary area of concern is governance of the transaction at the *organizational-level*, we move to property rights theory where we emphasize study of more macro-phenomena at the *institutional level*, beginning with Libecap's (1989) lucidly written research book.

## **APPENDIX: Cooperative Game and Mutual (Sunk-cost) Commitment**

Game theory is the analysis of rational behaviors in situations involving interdependent outcomes when the firm's payoff depends on what it does and what the other firm does (see Dixit and Nalebuff, 1991; McMillan, 1992; and Rasmusen, 1989). Since many strategic decisions involve interdependent outcomes, game-theoretic analysis can be applied, for example, to the study of vertical supplier-buyer relationships. In game-theoretic models each firm's optimal action depends on what the firm believes its counterpart will do. In other words, the game-theoretic analysis requires assumptions about the counterpart's rationality, and the counterpart's belief about the firm's own rationality.

The prisoners' dilemma game provides a powerful metaphor for a fundamental conflict that arises in business situations involving vertical interdependence (Saloner, 1991; Scherer and Ross, 1990). In the prisoners' dilemma what is best for the individual firm is to maximize its own economic profit and this "individual rationality" is detrimental to group performance. The "collective rationality" is for both firms to cooperate and obtain a higher group payoff, but the "individual rationality" is for each firm to play their dominant strategy, given the current payoff matrix. No matter how much they preach the importance of the group (common) good, there is always the possibility that the poor group outcome will be the dominant strategy equilibrium as predicted in the prisoners' dilemma game.

One way out of the prisoners' dilemma occurs when the players take steps that change the payoff matrix. Paradoxically, worsening some of one's own payoff possibilities may improve the likely outcome of the game (Schelling, 1960). Consider the case of an IT investment between Wal-mart and P&G. The best strategies for the collective good are that

both firms cooperate. While this mutual cooperation is collectively rational, it is unfortunately not individually rational in terms of individual firm profitability. Thus, in Payoff Matrix 1 we have an example of the Prisoners' dilemma situation where the Nash equilibrium point **(91, 91)** is predicted when each firm plays their dominant strategy of behaving opportunistically.

Payoff Matrix 1		Wal-mart	
		Cooperation	ppportunism
P&G	Cooperation	(112, 112)	8, 123)
	Oppportunism	123, 58)	<b>1, 91)</b>

Let's suppose P&G posts an economic bond (e.g., investing in a relation-specific IT system) that P&G would lose if P&G defects from the joint profit-maximizing collaboration. In effect, this action unilaterally lowers the payoff associated with an opportunistic behavior by P&G (i.e., from 123 to -28, and from 91 to -51, respectively below). Thus, P&G eliminates the attractiveness of defecting from the cooperative solution. Such a voluntary agreement is considered self-enforcing because third-party enforcement is not relied upon (Telser, 1980). As a result, Cooperation is now P&G's dominant strategy.

Payoff Matrix 2		Wal-mart	
		Cooperation	Oppportunism
P&G	Cooperation	(112, 112)	<b>8, 123)</b>
	Oppportunism	(-28, 58)	(-51, 91)

If Wal-mart, in a myopic manner, decides not to make some form of commitment to P&G, then Payoff Matrix 2 will not become the long-run equilibrium. P&G would eventually abandon their commitment to Wal-mart, and the situation would return to Payoff Matrix 1.

However, if Wal-mart wisely emulates P&G's action (i.e., mutually commits to a relation-specific IT with P&G) then this action further transforms the situation to Payoff Matrix 3 (as we see below) to encourage mutual commitments to cooperation that increases both firms' payoffs to **(112, 112)**. This example illustrates that firms involved in interdependent outcomes should seek both to give and receive mutual (sunk-cost) commitments that facilitate ongoing relationships and adaptation (Williamson, 1985). Thus, (sunk-cost) commitments are the dynamics of good strategy (Ghemawat, 1991).

Payoff Matrix 3		Wal-mart	
		Cooperation	Opportunism
P&G	Cooperation	<b>(112, 112)</b>	8, -28)
	Opportunism	(-28, 58)	(-51, -51)

It should be noted here that this game-theoretic example provides a "reconstructed logic" (Kaplan, 1964) of the IT alliance between P&G and Wal-mart. Whether this reasoning was the "logic-in-use" (Kaplan, 1964) of the managers of either or both of these firms is questionable. It is worth emphasizing that this game-theoretic reconstruction enables us to understand more fully the theoretical soundness of mutual commitments to support economic exchange.

### **Chapter 3 Property Rights Theory**

Coase (1960) initiated a flurry of property rights research that perhaps reached its peak with Alchian and Demsetz (1973). Barzel (1989) and Eggertsson (1990) provide useful discussions of the early property rights research literature. Much of this early property rights literature (with Demsetz, 1967 serving as an exemplar of the neoclassical economics tradition) was quite optimistic about the evolution of property rights toward economic efficiency. Three important criteria for efficiency of property rights are: (1) universality – all scarce resources are owned by someone; (2) exclusivity --- property rights are exclusive rights; and (3) transferability --- to ensure that resources can be allocated from low to high yield uses. In Demsetz' (1967) neoclassical economics framework all three criteria are in place (in the long-run).

In some sense, Libecap (1989), and especially North (1990), can be understood as providing historical accounts that challenge this earlier optimistic view of an inevitable evolution of property rights toward economic efficiency. The awarding of a Nobel Prize in economics to Douglass North suggests that, at the least, part of the economics profession has (implicitly) accepted the fact that the evolution of institutional environment change toward economic efficiency often fails.

Students studying the economics of organization should take note that changes in theoretical views do take place. However, in order to make headway, you need to come prepared with the facts along with an analytical approach (and often a tough skin) to handle the almost inevitable initial resistance by others to new ideas that aim to overturn the conventional wisdom.

We begin this chapter on property rights with Libecap's *Contracting for Property Rights*. Libecap (1989) provides substantive research concerning the way property rights are formed. Libecap's (1989) research book is a synthesis of theory and history, which emphasizes the complexities of property rights formation. Libecap (1989), in my judgment, is one of the best books in the property rights research literature. This research book is a major contribution both to the theory of property rights and to our understanding of economic history. In particular, we learn from Libecap (1989) that distributional conflicts present political risks to politicians, giving these politicians incentives to propose regulations that do not seriously upset status quo rankings and that offer only limited relief from property rights economics inefficiencies due to common pool resource losses. Similar incentives and vested interests exist for regulatory agencies.

North (1990) applies his theories of the interplay between institutional evolution and political and economic organization to a range of historical examples, including the development of management structures, insurance, and financial markets. North (1990) offers a broad perspective on how institutions persist and change. In particular, North (1990) is concerned as much with explaining the evolution of institutional frameworks that induce economic stagnation and decline as with accounting for the successes.

Eggertsson (1990) emphasizes the variety of organizational forms and institutional arrangements that we observe in practice. Eggertsson's (1990) approach to explain such variety is to seek a new synthesis of neoclassical economic theory and institutional theory. As Eggertsson (1990) views the research literature, three important levels are identified. At the first level, the structure of property rights and forms of organization are explicitly modeled but are treated as exogenous. At the second level, organization form is endogenous, but the fundamental structure of property rights remains endogenous. At the third level, attempts are made to consider

both social and political rules, and the structure of political institutions as endogenous in a positive transaction costs world. Eggertsson (1990) organizes his book on the basis of these three levels of analysis.

Barzel (1989), in the tradition of Coase (1960), provides a unified structure to analyze exchange, the formation of property rights, and organization. Barzel (1989) emphasizes that because of the costliness of measuring accurately all of an asset's attributes, rights are never fully delineated and property is consequently in danger of appropriation by others due to adverse selection, free-riding behavior, and shirking, among other reasons.

Hart (1990) argues that contractual incompleteness and control are two concepts that can be brought together to understand a number of economic institutions and arrangements. Hart (1990) focuses on understanding firms, and understanding financial structures. For the purposes of the current book, we will focus on the first half of Hart (1990) on understanding firms. Here, Hart (1990) focuses on some fundamental questions: (a) What does "ownership" mean? (b) What determines the boundaries of the firm? (c) What are the economic implications of contractual incompleteness? (d) What are the roles of non-human assets and the nature of authority?

In the 1990s, modern property rights theory (which provides more formalized mathematical models) has gained momentum in organizational economics, and Hart's (1995) work is an exemplar of this modern property rights framework. With the increasing importance of intellectual property rights in our current information age, (both early and modern) property rights theory predictably will receive greater attention in strategic management, and may prove to spur a growth area for research in the years ahead.



Libecap, Gary D. (1989). Contracting for Property Rights. Cambridge: Cambridge University Press.

How do institutions evolve in response to individual economic incentives, strategies, and choices? Libecap (1989) emphasizes that property rights matter.<sup>6</sup> Property rights provide the basic economic incentive system that shapes resource allocation. What has been largely missing is why property rights take the form that they do. Libecap (1989) argues that property rights are formed and enforced by political entities, and that property rights reflect the conflicting economic interests and bargaining strength of those affected. Moreover, because today's choices are constrained by yesterday's decisions, history matters.

Property rights are the social institutions that define or delimit the range of privileges granted to individuals to specific resources, such as parcels of land or water. Private ownership of these resources may involve a variety of property rights, including the right to exclude non-owners from access, the right to appropriate the stream of economic rents from use of and investments in the resource, and the rights to sell or otherwise transfer the resource to others. Property rights institutions range from formal arrangements, including constitutional provisions, statutes, and judicial rulings, to informal conventions and customs regarding the allocations and uses of property. Such institutions critically affect decision-making regarding resource use, and hence, affect economic behavior and economic performance.

Because of the huge advantages of secure property rights, economic decision-makers often are hypothesized to adopt, or to modify, property rights in order to mitigate the economic losses of the common pool, as soon as the private benefits of so doing outweigh the private costs.

Forces that drive the adjustments in property rights include new market prices, and production possibilities to which old arrangements are poorly attuned (Demsetz, 1988, 1995). Davis and North are explicit in the argument: "It is the possibility of profits that cannot be captured within the existing arrangement structure that leads to the formation of new (or the mutation of old) institutional arrangements" (1971: 39).

Despite these optimistic assertions in the (neoclassical) property rights literature, the actual process by which property institutions change, and whether the changes represent an efficient economic solution to a particular social problem, have received much less attention. North (1981: 6) notes that: "But the fact that growth has been more exceptional than stagnation or decline suggests that efficient property rights are unusual in history."

Libecap (1989) argues that because certain property rights arrangements can reduce transactions costs in exchange and production, and encourage (sunk cost) investments to promote overall economic growth, such property rights have public goods aspects. As with all public goods, though, there are economic hazards in attempting to change property rights. For example, there may be shirking and non-cooperative behavior among the bargaining parties that will affect the institutions that can be established. In bargaining over creating or modifying property rights, the positions taken by the various bargaining parties, including private claimants, bureaucrats, and politicians, will be molded by their *private* expected gains, as well as by the actions of the other parties.

Libecap (1989) emphasizes that property rights institutions are determined through the political process, involving either negotiations among immediate group members or the lobbying

---

<sup>6</sup> Seminal works in classical property rights theory include: Alchian (1965), Alchian and Demsetz (1973),

activities that take place at higher levels of government. The political process of defining and enforcing property rights can be divisive because of the distributional implications of different property rights allocations. If influential parties cannot be sufficiently compensated through share adjustments in the political process to win their support, beneficial institutional change (even as modified through contracting concessions) may not occur, and the potential economic gains fostered by the proposed arrangement will be forgone.

Even though society would be better off with the public goods provided by the new property rights, the distributional implications lead influential parties to oppose institutional change. In principle, it is possible to construct a side payment scheme that would compensate those who otherwise would oppose a desirable change in property rights. But in practice, devising perfectly compensating side-payments to bring agreement encounters formidable obstacles, including questions of who would receive side-payments, of who should pay, what should be the size of the compensation, and what should be the form of compensation. Libecap (1989) argues that distributional conflicts, and efforts to address such conflicts, can block institutional change or so influence the property rights arrangement that ultimately emerges that the institutional change bears little resemblance to that which was initially proposed.

The roles of time and precedent suggest that there may be historical path dependences for institutional change. Past property rights decisions serve to limit the menu of possible institutional solutions to varying economic problems. Libecap (1989) states that recent historical investigation suggests a less optimistic view of property rights change is in order. This conclusion is based on examination of the role of interest groups and conflicts among these groups over the distributional effects of property law and government regulation.

---

Barzel (1989), Cheung (1969), Coase (1960), Demsetz (1967), and Furubotn and Pejovich (1972).

Analytical Framework. Libecap (1989) notes that the nature in which property rights are defined and enforced, fundamentally impacts the performance of an economy for at least two reasons. First, by assigning ownership to valuable resources and by designating who bears the economic rewards and costs of resource-use decisions, property rights institutions structure incentives for economic behavior within the society. Second, by allocating decision-making authority, the prevailing property rights arrangement determines who are the key actors in the economic system.

In contracting over proposed property rights, the bargaining positions taken by the various parties depend upon how these parties view their economic welfare under the new arrangement relative to the status quo. Estimates of the likely net economic gains or losses from institutional change faced by each party require an evaluation of the overall productive possibilities with the new property rights arrangement and the distribution of economic rents it authorizes. The bargaining parties must see their economic welfare improved, or at least made no worse off, in order for them to support institutional change, and each party has an incentive to seek as large a share of economic rents under the new arrangement as possible. This competition for the range of economic opportunities made possible by changes in property rights is costly to society. Competition among the contracting parties uses resources, and such competition leads to changes in the definition and assignment of property rights that affect the nature and size of aggregate economic benefits that are possible. The side-payment schemes reached through the political process may be too incomplete to resolve the distributional conflicts needed for more than minimal institutional change to occur at any time.

Primary motivations for contracting for property rights are the aggregate (common pool) losses that arise under conditions of poorly defined property rights (e.g., open fisheries, oil field dissipation, and so forth). In these circumstances resource values fall for several reasons. First, because property rights to the resource are not assigned, individuals in their production decisions do not have to consider the full social costs of their activities. Individuals use the resource too rapidly at any time, relative to interest rate and price projections. Further, competitive pressures under conditions of poorly defined property rights encourage short-time horizons in production. The economic incentive to invest (e.g., in new technology) is reduced because investors cannot anticipate that they will capture any of the resulting economic returns due to insecure property rights.

Second, resource values fall because exchange and reallocation of the resource to higher-valued uses becomes more costly and less effective if property rights are absent. Demsetz (1967) argues that an assignment of property rights is a prerequisite before decentralized price-making markets can form to define asset prices. Well-defined asset prices are needed to reflect underlying demand and supply conditions, and to facilitate socially valuable exchange among economic agents. Without the more complete market signals possible when property rights are well defined, resources may not flow smoothly to higher-valued uses as economic conditions change. Whether or not the more complete defining of property rights is socially beneficial depends on the magnitude of common pool losses, the nature of contracting costs to resolve such losses, and the economic costs of defining and enforcing property rights (Coase, 1960).

In Libecap's (1989) analytical framework, pressures to change existing property rights can emerge from the following factors:

- Shifts in relative prices;

- Changes in production and enforcement technology; and
- Shifts in preferences and other political parameters.

A number of implications can be drawn from Libecap's (1989) analytical framework:

- (1) All things being equal, the greater the size of the anticipated aggregate economic benefits of institutional change (the greater the economic losses of the common pool), the more likely new property rights will be sought and adopted because it is more likely that a politically acceptable share arrangement can be devised by politicians to make enough influential parties better off so that institutional change can proceed;
- (2) The larger the number of competing interest groups, the more likely distributional conflicts will block or delay institutional change because the greater the number of competing interest groups with a stake in the new definition of property rights, the more claims that must be addressed by politicians in building a consensus for institutional change;
- (3) The greater the heterogeneity of competing interest groups, the more likely distributional conflicts will block or delay institutional change. Important differences across the parties in information regarding the resource, as well as in production costs, size, wealth, and political experience, will make the formation of winning political coalitions, and a consensus on the proposed assignment or adjustment of property rights, more difficult;
- (4) Distributional conflicts will be intensified if there are known serious information asymmetries among the competing parties regarding the evaluation of individual claims. These distributional conflicts will occur quite aside from any strategic bargaining efforts if private estimates of the economic value of current property rights, and of potential economic losses from the new system, cannot be conveyed easily or credibly to politicians and the other bargaining parties; and
- (5) The greater the concentration of wealth under the proposed property rights allocation, the greater the likelihood of political opposition and the less likely institutional change will be adopted without modification by politicians. In these circumstances, enough influential parties may see their economic welfare made worse, or at least not improved, by the change that political support for such change does not materialize.

*Contracting for the Unitization of Oil Fields.* Libecap (1989) observes that since the first discovery of petroleum in the United States in 1859, oil production has been plagued by serious common pool losses. These common-pool losses arise as numerous firms compete for

migratory oil lodged in sub-surface reservoirs. Under the common rule law of capture, private property rights to oil are assigned only upon extraction. For each of the firms on a reservoir, a plan of dense-well drilling and rapid production allows the firm to drain oil from its neighbors and to take advantage of the low extraction costs that exist early in oil field development. In new flush oil fields, sub-surface pressures are sufficient to expel oil without costly pumping or injection of water or natural gas into the reservoir to drive oil to the surface.

Libecap (1989) notes that under these conditions, when there are multiple firms on a reservoir, each firm has an economic incentive to drill competitively and to drain to increase its share of oil field economic rents, even though these individual actions lead to aggregate common pool losses. Economic rents are dissipated as capital costs are driven up with the drilling of excessive numbers of wells (more than geological conditions require or price and interest rate projections warrant) and with the construction of surface storage, where the oil can be held safe from drainage by other firms. Unfortunately, once in surface storage, oil is vulnerable to fire, evaporation, and spoiling. Rapid extraction also increases production costs as sub-surface pressures are vented prematurely, forcing the early adoption of pumps and injection wells. Total oil recovery falls as pressures decline because oil becomes trapped in surrounding formations, retrievable only at high extraction costs. Finally, economic rents are dissipated as production patterns diverge from those that would maximize the economic value of output over time. Some estimates indicated that oil recovery rates of only 20 to 25 percent occurs with competitive extraction, whereas recovery rates of 85 to 90 percent were thought possible with controlled withdrawal.

A complete solution to the common pool problem is oil field-wide unitization. Under unitization, production rights are delegated through negotiations to a single firm, the unit

operator, with net revenues apportioned among all parties on the field (including those that would otherwise be producing). As the only producer on the field and a residual profit claimant, the unit operator has an economic incentive to maximize field rents. Accordingly, unitization results in important economic gains: a time stream of output that more closely approximates the rent-maximizing pattern, increased oil recovery (two to five times greater than unconstrained production), and reduced wells and other capital costs. Despite these reasons for mitigating the substantial losses involved in common pool crude oil production, complete field-wide unitization had not been widespread. As late as 1975 only 38 percent of Oklahoma production and 20 percent of Texas production came from field-wide units.

Libecap (1989) argues that the key issue in blocking agreement on the voluntary unitization of oil fields is the distributional conflict over the share formula to divide the net proceeds of unit production among the various contractual parties. Uncertainties and information asymmetries regarding the economic valuation of individual firms oil leases, which are the basis for unit shares, are important contributors to the disagreements that block unitization, even in the presence of large and uncontroversial aggregate economic gains from unit formation. In share negotiations two serious problems arise. First, unitization contracts must assign, once-and-for-all, shares at the time the contract is completed. This assignment is needed because in reservoir dynamics after unitization, it is impossible to link unit production to particular leases, which would be necessary for adjusting shares. A second problem in unitization contracting is general uncertainty and asymmetrical information regarding relative pre-unitization lease values, which determine unit shares. These serious contractual problems block agreement on lease value estimates and proposed shares in unit economic rents.



Besides the information issues, small lease owners were given preferential drilling permits by regulatory authorities under pro-rationing controls adopted by states in the absence of widespread unitization. Differences in lease value estimates can block consensus on any side-payments to draw potential holdouts into agreement. Under unanimity voting rules, small firms could delay or block the formation of field-wide units. The empirical evidence that Libecap (1989) presents supports the notion that as field-wide primary production nears an end, unitization agreements become more likely. By that time, information asymmetries among the firms become less important as all leases near primary depletion.

The failure of unitization to be widespread, despite significant aggregate economic gains from unitizing oil production, is another example of how distributional conflicts over rental shares can limit the adoption of property rights to increase economic efficiency. *The analysis presented by Libecap (1989) suggests that swift institutional responses to common pool losses to promote more rational resource use and greater economic growth cannot be taken for granted. Distributional conflicts inherent in any new property rights arrangement can block, or critically constrain, the institutions that can be adopted.* More attention accordingly, should be directed to the distributional implications of property rights arrangements, to the identity and preferences of the various bargaining parties, and to the nature of the side payment schemes adopted. And, perhaps even more importantly, attention should be directed to the history of past political agreements if the observed variations in property rights and associated economic and strategic behaviors are to be more fully understood.<sup>7</sup>

---

<sup>7</sup> Libecap and Wiggins (1985) provide empirical evidence of the influence of private contractual failure on regulation for the case of oil field unitization. Kim and Mahoney (2002) provide a fairly comprehensive collection of references on the property rights approach and provide resource-based and property rights perspectives concerning oil field unitization. Finally, elements of the oil field unitization

Libecap (1989) provides an exemplar for students studying the economics of organization on the use of case studies to build up and support a theoretical argument. Libecap (1989), in my judgment, convincingly shows that the assertion that property rights will naturally move toward economic efficiency is frequently glib and inaccurate.

Now that we have studied Libecap (1989), we next examine the work of North (1990). Early in his career (e.g., Davis and North, 1971), North held an (overly) optimistic view about the evolution of property rights toward economic value creation. In contrast, North (1990) emphasizes the persistence of inefficient property rights regimes throughout economic history to provide a main case explanation for why the whole world is not economically developed. The objective of North's (1990) research book is to provide an analytical framework to integrate institutional analysis into economics and economic history. North (1990) provides us with a new understanding of historical change.

In this research book, North (1990) examines the nature of institutions and the consequences of institutions for economic and societal performance. North (1990) then outlines a theory of institutional change not only to provide a framework for economic history, but also to explain how the past influences the present and future, the way incremental institutional change affects the choice set of decision-makers at a moment in time, and the nature of path dependencies. The primary objective of this research book is to achieve an understanding of the differential performance of economies through time.

North (1990) ties together the threads and illustrates the relationships among institutions, transaction costs, and transformation (production) costs. North (1990) then explores

---

case discussed in this chapter provide insight on the conflicts between Kuwait and Iraq that lead to the

organizations and the way that they interact with institutions. North (1990) argues that the nature of incremental institutional change together with the imperfect way by which decision-makers interpret their environment and make choices accounts for path dependencies, and makes history relevant.

---

Persian Gulf War of 1990-1991 (for some details, see Milgrom and Roberts, 1992: 296).

North, Douglass C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.

North (1990) asks: What combination of institutions best permits capturing the economic gains from trade? Institutions are defined as any constraint humans devise to shape their interactions and organizations, created to take advantage of the opportunities presented by institutions in shaping the development of economies. The importance of institutions arises from the costliness of measuring what is valuable, from protecting rights, and from policing and enforcing agreements.

North (1990) emphasizes that history matters. History matters not just because we can learn from the past, but also because the present and the future are connected to the past by the continuity of a society's institutions. Today's decisions and tomorrow's choices are shaped by the past. And the past can only be made intelligible as a story of institutional evolution.

For North (1990), the central focus is on the problem of human cooperation — specifically, the cooperation that permits economies to capture the economic gains from specialization and trade. The evolution of institutions that create a hospitable environment for cooperative solutions to complex exchange provides for economic growth.

North (1990) argues that institutions reduce uncertainty by providing a structure to everyday life. Institutions are a guide to human interactions, and these institutions define and limit the set of choices of individuals. Institutions include any form of constraint that humans devise to shape human interaction. Are institutions formal or informal? Institutions can be either, and North (1990) considers both formal constraints — such as rules that humans devise — and informal constraints — such as conventions and codes of behavior. Institutions may be

created, as was the United States Constitution; or institutions may evolve over time, as does the common law. An essential part of the functioning of institutions is the costliness of ascertaining violations and the severity of punishment.

North (1990) makes an important distinction between institutions and organizations. Organizations include political bodies (e.g., political parties, trade unions, family farms, cooperatives), social bodies (e.g., churches, clubs, athletic associations), and educational bodies (e.g., schools, universities, vocal training centers). Organizations are groups of individuals bound by some common purpose to achieve objectives. Modeling organizations requires analyzing governance structures, organizational capabilities, and how learning-by-doing will determine the organization's success over time (Oliver, 1997). The institutional framework fundamentally influences both what organizations come into existence and how organizations evolve. In turn, organizations influence how the institutional framework evolves.

North (1990) emphasizes that institutions are a creation of humans. North (1990) suggests that integrating individual choices with the constraints that institutions impose on choice sets is a major step toward unifying social science research. The major role of institutions in society is to reduce uncertainty by establishing a stable (but not necessarily efficient) structure to human interactions. Although formal rules may change overnight as the result of political or judicial decisions, informal constraints embodied in customs, traditions, and codes of conduct are much more impervious to deliberate policies. These cultural constraints not only connect the past with the present and future, but also provide us with a key to explaining the path of historical change.

*North (1990) maintains that the central puzzle of human history is to account for the widely divergent paths of historical change.* North (1990) notes that although we do observe

some convergence among leading industrial nations that trade with each other, an overwhelming feature of the last ten millennia is that we have evolved into radically different religions, ethnic, cultural, political, and economic societies. Furthermore, the economic gap between rich and poor nations, between developed and underdeveloped nations, is as wide today as it ever was, and perhaps a great deal wider than ever before.

North (1990) then asks: What accounts for societies experiencing long-run stagnation or an absolute decline in economic well-being? North and Thomas (1973) make institutions the determinant of economic performance and relative price changes were made the source of institutional change. North and Thomas (1973) provide an essentially efficiency-based explanation; changes in relative prices create economic incentives to construct more efficient institutions. North (1981), however, abandons the efficiency view of institutions. *Rulers devised property rights in their own vested interests and transaction costs resulted in typically inefficient property rights prevailing. As a result, it was possible to account for the widespread existence of property rights throughout history (and in the present) that did not produce economic growth.*

North (1990) argues that institutions highly influence the opportunities in a society. Organizations are created to take advantage of those opportunities, and, as the organizations evolve, organizations alter the institutions. The resultant path of institutional change is shaped by: (1) the lock-in that comes from the tightly coupled relationship between institutions and organizations that have evolved as a consequence of the economic incentive structure provided by those institutions; and (2) the feedback process by which humans perceive, and react to, changes in the (subjective) opportunity set.

Actors frequently must act on incomplete information and process the information that they do receive through mental constructs, which can result in persistently inefficient paths. Transaction costs in political and economic markets make for inefficient property rights, but the imperfect subjective models of the actors as they attempt to understand the complexities of the problems they confront can lead to the persistence of inefficient property rights.

North (1990) states that there is a persistent tension in the social sciences between the theories we construct and the evidence we compile about human interaction in the world around us. This tension is most striking in economics, where the contrast between the logical implications of neoclassical microeconomic theory and the performance of economies (however defined and measured) is startling. *North (1990) argues that the coercive power of the State has been employed throughout most of history in ways that have stymied economic growth.*

North (1990) maintains that the traditional behavioral assumptions of orthodox microeconomic theory have prevented economists from coming to grips with some fundamental issues that have impeded progress in the social sciences. In particular, North (1990) argues that the motivation of actors is more complicated (and their preferences less stable) than assumed in the received wisdom. Further, microeconomic theory implicitly assumes that actors possess cognitive systems that provide *true* models of the worlds about which they make choices. North (1990) insists that this implicit assumption is patently wrong for most of the important problems with which institutional economics and organizational economics are concerned. Individuals make choices based on subjectively derived models that diverge among individuals and the information of actors is so incomplete that in most cases these divergent subjective models show no tendency to converge. Only when we understand these modifications in the behaviors of the

actors can we hope to make sense out of the existence and structure of institutions, and to explain the direction of institutional change.

North (1990) argues that institutional analysis requires that we delve into two particular aspects of human behavior: (1) motivation and (2) deciphering the environment. Many cases are not simply of wealth-maximizing behavior, but of altruism and of self-imposed constraints, which radically change the outcomes with respect to the choices that people actually make. Similarly, we find that people decipher the environment by processing information through pre-existing mental constructs through which they understand the environment and solve the problems they confront.

North (1990) notes that the works of Simon (1982) capture the essence of why the subjective and incomplete processing of information plays a crucial role in decision-making. Simon's (1982) work is useful for accounting for ideology, based upon subjective perceptions of reality, playing a major role in humans' choices. Simon's (1982) work brings into play the complexity and incompleteness of our information, and the fumbling efforts we make to decipher information. North (1990) concludes that the regularized interactions that we call institutions may be inadequate to deal with the economic problems at hand.

Culture can be defined as the transmission from one generation to the next, the teaching and replication of knowledge, values, and other factors that influence behaviors. North (1990) argues that culture provides a language-based conceptual framework for encoding and interpreting the information that the senses are presenting to our brain. Importantly, the cultural filter provides continuity and stability. Order is the result of a dense social network where people have an intimate understanding of each other. In the short-run, culture defines the way



individuals process and utilize information, and hence may affect the way informal constraints get specified. Conventions are culture specific, as indeed are informal rules.

Formal Constraints. North (1990) observes that formal rules can complement and increase the effectiveness of informal rules. Formal rules also may be enacted to modify, revise, or replace informal constraints. Formal rules include political (and judicial) rules, economic rules, and contracts. Economic rules define property rights and as a crude approximation, economic rules are derived from economic self-interest. Property rights are specified and enforced by political decision-making, but the structure of economic interests will also influence the political structure. Indeed, there is a substantial amount of property rights literature that looks on the development of property rights as a simple function of changes in economic costs and economic benefits. North (1990) argues that this simplified approach needs modification to account for the obvious persistence of inefficient property rights.

Enforcement. North (1990) argues that the inability of societies to develop effective, low-cost enforcement of contracts is the most important source of both historical stagnation and contemporary under-development in the Third World. In developed countries, effective judicial systems include well-specified bodies of law and agents such as lawyers, arbitrators, and mediators, and one has some confidence that the merits of a case rather than private payoffs will decisively influence outcomes. In contrast, enforcement in the Third World economies is uncertain not only because of ambiguity of legal doctrine (a measurement cost), but also because of uncertainty with respect to behavior of the judicial system.

Institutions. North (1990) observes that it takes resources to define and protect property rights, and to enforce agreements. Institutions together with the technology employed determine those transaction costs. It takes resources to transform inputs of land, labor, and capital into the

output of goods and services, and that transformation is a function not only of the technology employed, but of the institutions as well. Therefore, institutions play a key role in the costs of production. The interplay between techniques, institutions, transformation costs, and transaction costs make clear that the relationships among them are complex.

North (1990) submits that contrasting the institutional framework in countries such as the United States, England, France, Germany, and Japan with Third World countries makes clear that the institutional framework is the critical success factor of economies, both cross-sectionally as well as through time. North (1990) further argues that the institutional framework shapes the direction of the acquisition of knowledge and capabilities, and that direction will be the decisive factor for the long-run development of that society. Path dependence is the key to an analytical understanding of long-run change in property rights. Property rights and economic incentives are the underlying determinants of economic performance. Bringing property rights and economic incentives to the foreground focuses attention where it belongs, on the key success factors for the economic performance of societies. One gets *efficient* institutions by a polity that has built-in economic incentives to create and enforce efficient property rights.

North (1990) concludes that we need to know much more about culturally derived norms of behavior and how such norms of behavior interact with formal rules to get better answers to such issues. We are just beginning the serious study of institutions in organizational economics and strategic management. The promise is there. We may never have definitive answers to all our questions. But students in the next generation of research can do better in both institutional economics and organizational economics research, which will contribute greatly to the evolving science of organization.

Our third property rights book by Barzel (1989) is *An Economic Analysis of Property Rights*, and our fourth book is an over-looked classic by Eggertsson (1990) on *Economic Behavior and Organizations*. The Barzel (1989) book is complementary to Libecap (1989), and the Eggertsson (1990) book is especially complementary to North (1990). In fact, Eggertson notes his intellectual debt to Douglass North: “North’s vision that the economic approach, augmented by transaction costs and property rights, is a general tool for the study of society at all levels has inspired this book” (1990: xiv).

Barzel, Yoram (1989). Economic Analysis of Property Rights. Cambridge: Cambridge University Press.

Barzel (1989) notes that because transacting is costly, as an economic matter property rights are never fully delineated. Property rights of individuals over resources consist of the rights, or the powers, to consume, obtain income from, and alienate those resources. Obtaining income from and alienating resources require exchange; and exchange is the mutual ceding of rights. Legal rights, as a rule, enhance economic rights, but legal rights are neither necessary nor sufficient for the existence of the economic rights. The rights people have over resources (including themselves and other people) are not constant; they are a function of their own direct efforts at protection, of other people's capture attempts, and of government protection.

Barzel (1989) views the concept of property rights to be closely related to that of transaction costs. Transaction costs are defined as the economic costs associated with the transfer, capture, and protection of rights. When transaction costs are positive, rights to resources cannot be perfectly delineated. Exchange that otherwise would be attractive may be forsaken because of such exchange costs.

What underlies this costliness of transacting? What are the factors that prevent people from realizing the full economic value of their resources? Commodities have many attributes whose levels vary from one specimen of a commodity to another. The measurement of these levels is too costly to be comprehensive or entirely accurate. How difficult it is to obtain full information in the face of variability, fundamentally determines how difficult it is to delineate rights. *Because it is costly to measure commodities fully, the potential of wealth capture is present in every exchange.* The opportunity for wealth capture is equivalent to finding property

in the public domain; in every exchange, then, some wealth spills over in the public domain, and individuals spend resources to capture this economic wealth. Whereas people always expect to gain from exchange, they also always spend resources on the capture of economic wealth. Individuals maximize their (expected) net gains, the gains from exchange as conventionally perceived net of the economic costs of effecting exchange.

The sale of cherries illustrates the phenomenon of wealth capture. Obvious problems of information present themselves when cherries are exchanged. Customers must spend resources in order to determine whether a store's cherries are worth buying, and in order to determine which particular cherries to buy. Storeowners who allow customers to pick and choose cannot easily prevent these customers from eating cherries after these customers decided whether or not to buy the cherries, nor can storeowners prevent customers' careless handling of cherries. Indeed, the process of picking and choosing itself allows wealth capture in the form of excess choosing. Some of the cherries' attributes, then, are placed in the public domain. The high cost of information results in transaction costs: economic costs that would not arise were the owner and the consumer of cherries the same person. If information about the cherries were costless, their initial owner would not have to relinquish any rights; and pilfering, damage, and excess choosing would be avoided. In business reality, such public domain problems are unavoidable; people can take steps, however, to reduce the associated economic losses.

Contracts govern the exchange of property rights and are central to the study of such rights. The exchange value of a resource is a function of the gross income the resource can generate, and of the transaction costs of measuring and policing its exchange. These economic costs also determine the pattern and the degree of ownership. The ownership of a resource's

attributes is expected to gravitate into the hands of those people who are most inclined to affect the income flows that the attributes can generate.

Barzel (1989) maintains that the property rights transaction costs model can generate a better understanding of the allocation of resources, and of the interaction of this allocation with economic organization. The research literature that assumes that the economic costs of transactions are zero and that all property rights are perfectly well delineated is incapable of dealing with a vast array of actual observed practices. Particularly glaring is the inability of such an approach to explain why exchange parties would ever impose restrictions on each other. The property rights approach is capable of addressing such issues, and we continue our property rights study with Eggertsson (1990).

Eggertsson (1990) considers the costs of transacting and the allocation of resources; transaction costs and efficiency; the quality dimensions of goods and the costs of measurement, property rights and their dimensions; the partitioning of property rights; property rights and contract theory; the emergence of property rights; competition and the costs of alternative economic organizations, and economic outcomes. This research book provides a clear structure to and balanced overview of the property rights literature. Eggertsson (1990) provides a mature yet compact presentation of property rights research.

Eggertsson, Thrainn (1990). Economic Behavior and Institutions. Cambridge: Cambridge University Press.

Eggertsson (1990) observes that organizations and institutions are not invariant; organizations and institutions vary with time and location, with political arrangements and structures of property rights, with technologies employed, and with physical qualities of resources and services that are exchanged. In fact, production involves not only the physical transformation of inputs into outputs but also the transfer of property rights between the owners of resources, and labor services.

Eggertsson (1990) refers to the rights of individuals to use resources as *property rights*. A system of property rights is a method of assigning to particular individuals the “authority” to select, for specific goods, any use from an un-prohibited class of uses. The rights of individuals to the use of resources (i.e., property rights) in any society are supported by the forces of etiquette, social custom, ostracism, and are also supported by formal laws that are backed up by the States’ power of coercion.

It is common to distinguish three categories of property rights: First, there are the rights to use a resource including the right to transform physically a resource. Second, there is the right to earn income from a resource and contract over the terms with other individuals. Third, there is the right to transfer permanently to another party ownership over a resource — i.e., to alienate or sell a resource.

The enforcement of property rights includes excluding others from the use of scarce resources. Exclusive ownership calls for costly measurement and delineation of resources, and enforcement of ownership rights. The economic value of exclusive ownership rights depends, *ceteris paribus*, on the costs of enforcing those rights — that is, the costs of excluding others,

which ultimately depends on coercion. The enforcement of exclusive rights is usually undertaken by both individual owners and by the State.

An economic problem arises when property rights over a valuable resource — e.g., the rights to the air over the factory and the neighborhood — have not been fully delineated. In fact, the dispute between the factory and the neighborhood community involves a struggle over access to a common property resource. Once ownership over the atmosphere is established, the economic problem can be resolved. In the real business world, we often find that rights to valuable resources are not fully delineated. Reasons for why these property rights are not fully delineated include: a weak State, high measurement costs relative to the economic value of a resource, rapid economic change, and struggles over the distribution of wealth.

Property rights to a resource are often *partitioned*. For example, in the case of land, person A and person B may possess the right to grow wheat on the land. Person C may possess the right to dump ashes on the land. Person D may possess the right to fly an airplane over the land. And, each of these rights may be transferable. In sum, private property rights to various partitioned uses of land are “owned” by different persons.

According to the so-called Coase (1960) theorem, the initial partitioning of property rights does not matter for the allocation of resources (ignoring wealth effects) when all rights are freely transferable, and the costs of transacting are zero. But when transaction costs are introduced, the role of the State can have a crucial effect on resource allocation. Negotiation costs and other transaction costs may block the re-assignment of rights, and the initial partitioning of property rights by the State may have important consequences for the output of an economy. Thus, the property rights approach is not complete without a theory of the State.



Eggertsson (1990) notes that the structure of a contract depends on the legal system, social customs, and the technical attributes of the resources involved in exchange. The more detailed the legal framework and the stronger the ties of custom and social control, the less specific are the written contracts. The State by using its police power and the courts, assists private individuals in enforcing legitimate contracts, and thus lowers the costs of exchange, particularly when the State uses its power to enforce contracts in a *systematic and predictable manner*. In a business world of positive transaction costs, the distribution of political power within a country and the institutional structure of its rule-making institutions, are critical success factors in economic development.

Demsetz offers an optimistic theory of property rights: “Property rights develop to internalize externalities when the gains of internalization become greater than the cost of internalization” (1967:350). *Eggertsson (1990) notes that characteristic of this optimistic view, the formulation of decision-making with regard to property rights, is solely in terms of private benefits and private costs. The theory does not deal with the free-riding problems that plague group decision, nor is there an attempt to model political processes.* However, as Libecap (1989) demonstrated earlier in this chapter, the State does *not* always act to minimize costs and maximize economic value. In particular, the state governments of Texas and Oklahoma failed to design rules that encouraged the unitization of oil fields.

Eggertsson (1990) argues that a rudimentary knowledge of economic history or modern economic systems rules out Demsetz’ (1967) optimistic model as a general theory. One of the first steps to modify the optimistic model of property rights involves linking this model to the interest-group theory of legislation and government. Eggertsson (1990) refers to this extension of the optimistic model as the *interest-group theory of property rights*.

The interest-group theory of property rights takes the fundamental social and political institutions of the community as given, and seeks to explain the structure of property rights, in various industries, in terms of interactions between interest groups in the political market. Property rights, which serve the narrow self-interest of special-interest groups but cause substantial output losses to the community as a whole, typically are explained in terms of transactions costs, free-riding, and asymmetrical information. Eggertsson (1990) concludes (along with North, 1990) that there is overwhelming historical evidence to support the proposition that States typically do *not* supply structures of property rights that are appropriate for placing the economy close to the technical production frontier.

The first four books in this chapter have been in the “classical” property rights literature. We conclude this chapter with the modern (more formalized) property rights theory (e.g., Grossman and Hart, 1986; Hart and Moore, 1990) and the exemplar work of Hart (1995): *Firms, Contracts, and Financial Structure*. Hart (1989, 1995) focuses on the boundary and scope of the firm in the market economy and describes an incomplete contracting or “property rights” approach to both explain and predict firm-level vertical integration decisions. Hart (1995) emphasizes the meaning and importance of asset ownership.

Hart, Oliver (1995). Firms, Contracts, and Financial Structure. Oxford: Clarendon Press.

Hart (1995) provides a framework for thinking about firms and other kinds of economic institutions. The basic idea is that firms arise in situations where people cannot write complete contracts, and where the allocation of control is therefore important. Given that we will write an incomplete contract, it is clear that revisions and renegotiations will take place. In fact, the contract is seen as a suitable starting point for such renegotiations rather than specifying the final outcome. Hart (1995) submits that because contracts are incomplete, the ex post allocation of control matters. Indeed, these two ideas, contractual incompleteness, and the ex post allocation of control, can be used to understand a number of economic institutions.

Property rights theory focuses on how control rights are allocated in a contractual relationship when contracts are incomplete. Hart (1995) notes that in principal-agent theory it is supposed that it is costless to write a contract. An implication is that an optimal contract will be comprehensive in the sense that the optimal contract will stipulate each person's obligations in every conceivable eventuality and impose large economic penalties if anybody fails to live up to these obligations. Control issues are irrelevant in the principal-agent model since an optimal comprehensive contract will not be renegotiated.

Hart (1995) also observes that transaction costs theory comes closest to the framework of the modern property rights theory. However, although transaction costs theory puts a lot of emphasis on the economic costs of writing contracts and the consequent contractual incompleteness, less attention is paid to the idea that institutional arrangements are designed to allocate control rights among agents.

The Meaning of Ownership. Hart (1995) points out that scholars have written a great deal about why property rights are important, and in particular why it matters whether a machine, say, is privately owned or is common property. However, there has been less success in explaining why it matters *who* owns a piece of private property. To understand the difficulty, consider a situation where I want to use a machine initially owned by you. One possibility is for me to buy the machine from you; another possibility is for me to rent the machine from you. If contacting costs are zero, we can sign a rental agreement that is as effective as a change in ownership. In particular, the rental contract can specify exactly what I can do with the machine, when I can have access to it, what happens if the machine breaks down, what rights you have to use the machine, and so on. Given this possibility, however, it is unclear why changes in asset ownership ever need to take place.

In a business context where there are positive transaction costs, however, renting and owning are no longer economically equivalent. If contracts are incomplete, not all the uses of the machine will be specified in all possible eventualities. The economic question then arises: who chooses the unspecified uses? A reasonable approach is that the owner of the machine has this property right; that is, the owner has the residual rights of control over the machine. For example, if the machine breaks down or requires modification and the contract is silent about this contingency, the owner can decide how and when the machine is to be repaired or modified. It is now possible to understand why it might make sense for me to buy the machine from you rather than to rent the machine from you. If I own the machine, I will have all the residual rights of control. To put it another way, if the machine breaks down or needs to be modified, I can ensure that the machine is repaired or modified quickly, so that I can continue to use the machine

productively. Knowing this possibility, I will have a greater economic incentive to look after the machine, to learn to operate the machine properly, and to acquire other machines that create a synergy with this machine.

The Boundaries of the Firm. Once we recognize that contracts are incomplete and transaction costs are positive, then the boundaries of the firm matter for economic efficiency. Specifically, Hart (1995) argues that firm boundaries are chosen to allocate control rights optimally among the various parties to a transaction. A merger between firms with highly complementary assets enhances economic value. If two highly complementary firms have different owners, then neither owner has real control since neither can do anything without the other. It is better to give all the control rights to one of the owners through a merger.

Agency Theory Hart (1995) observes that neoclassical microeconomic theory ignores all economic incentive problems within the firm. Over the last twenty years or so, a branch of the organizational economics research literature --- principal-agent theory --- has developed that tries to rectify this neglect of an essential organizational economic problem. We will consider in more detail principal-agent theory in the next chapter. Hart (1995) argues that principal-agent theory leads to a richer and more realistic portrayal of firms, but that principal-agent theory leaves unresolved the basic issue of the determinants of firm boundaries.

Hart (1995) notes that there is now a vast research literature that analyzes the form of the optimal economic incentive scheme under specified circumstances. Moreover, the basic principal-agent problem described has been extended in a number of directions. Among other things, agency theorists have allowed for: repeated relationships, several agents, several principals, several dimensions of actions for the agent, career concerns, and reputation effects.

As a result of all this research, a rich set of results about optimal economic incentive schemes has been obtained. However, although these results can throw important light on the determinants of managerial compensation packages and on certain aspects of the organization of production, the agency approach does not pin down the *boundaries* of the firm (or say much about the internal organization of firms).

Hart (1995) points out that agency theory does not distinguish an optimal contract written by independent firms and internal transfers between divisions of a firm, and yet economically they are quite different. The principal-agent theory is consistent with there being many small, independent firms linked by optimal arm's length contracts; but this theory is also consistent with there being one large firm, consisting of a large number of divisions linked by optimal economic incentive contracts. Clearly, there is something missing from the agency theory of the firm (just as there is something missing from the neoclassical theory of the firm).

The Distinction Between Comprehensive and Incomplete Contracts. Hart (1995) argues that one important factor missing from the principal-agent view is the recognition that writing a (good) contract is itself costly (Coase, 1988; Williamson, 1985). Hart (1995) maintains that although the optimal contract in a standard principal-agent model will not be first-best (since it cannot be conditioned directly on variables like effort that are observed by only one party), the optimal contract in a standard principal-agent model will be comprehensive in the sense that the principal-agent model will specify all parties' obligations in all future states of the world, to the fullest extent possible. As a result, there will never be a need for the contractual parties to revise or renegotiate the contract as the future unfolds. The reason is that, if the contractual parties ever changed or added a contract clause, this change or addition could have been anticipated and built

into the original (comprehensive) contract. One would also not expect to see any legal disputes in a comprehensive contracting world. The reason is that, since a comprehensive contract precisely specifies everybody's obligations in every eventuality, the courts should simply enforce the contract as it stands in the event of a dispute.

The Sources of Transaction Costs. Hart (1995) notes that in business reality, contracts are not comprehensive, and are revised and renegotiated all the time. According to the transaction costs research literature, renegotiation is a consequence of three factors missing from the standard principal-agent model:

- In a complex and highly unpredictable business world it is difficult for people to *think* far ahead and to plan for all the various contingencies that may arise;
- Even if individual plans can be made, it is hard for the contracting parties to *negotiate* about these plans, not least because the contractual parties have to find a common language to describe states of the world and actions with respect to which prior experience may not provide much of a guide; and
- Even if the contractual parties can plan and negotiate about the future, it may be difficult for them to *write* their plans down in such a way that, in the event of a dispute, an outside authority --- a court, say --- can figure out what these plans mean and enforce these plans.

Hart (1995) concludes that as a result of these three contracting costs, the parties will write a contract that is incomplete. That is, the contract will contain gaps and missing provisions.

The Economic Implications of Contractual Incompleteness. Hart (1995) notes that as observed, an incomplete contract will be revised, and/or renegotiated as the future unfolds. In fact, given that the contractual parties can fill in the gaps as they go along, one may ask why contractual incompleteness matters. The reason is that the renegotiation process imposes several

transaction costs. Some of these costs are *ex post* costs incurred at the renegotiation itself, and others are *ex ante* costs incurred in anticipation of renegotiation.

First, the contractual parties may engage in a great deal of haggling over the terms of the revised contract. Argument about division of surplus serves no overall productive purpose and, to the extent that haggling is time-consuming and wastes resources, such haggling is inefficient. Second, there may be costly legal disputes because an incomplete contract will be ambiguous, and the contractual parties will look to the courts to resolve the ambiguity. Third, not only may the process of *ex post* bargaining be costly, but also to the extent that the contractual parties have asymmetric information, the contractual parties may fail to reach an efficient agreement.

Hart (1995) argues that if these three costs are high, it must be because there is something binding the partners together and making it difficult for them to switch at the re-contracting stage. The leading candidate for that “something” is an *ex ante relationship-specific investment*, that is, a prior strategic commitment, which creates economic value if the contractual parties’ economic relationship extends over time.

Hart (1995) maintains that once the existence of relationship-specific investments is recognized, it becomes apparent that there can be a third cost of contractual incompleteness that may dwarf the haggling and *ex post* inefficiency costs. Specifically, because contracts are incomplete, the contractual parties may be deterred from making the relationship-specific (sunk cost) investments that would be optimal in a “first-best” world. Given each contractual party’s fear that the other party will “hold it up” at the renegotiation stage, the contractual parties are likely to make investments that are relatively non-specific. Such decisions sacrifice some of the efficiency benefits of specialization, but, in a world of incomplete contracting, these efficiency



losses are more than offset by the security that a non-specific investment provides for each contractual party.

Hart (1995) asks: How would these costs change if the two independent, i.e., non-integrated firms merged and became a single firm? If there is less haggling and hold-up behavior in a merged firm (as transaction costs theory submits), it is important to provide reasons *why*. The modern property rights approach developed by Grossman and Hart (1986) and Hart and Moore (1990), the so-called Grossman-Hart-Moore (GHM) model, focuses on this *efficient boundaries* question.<sup>8</sup>

*The Property Rights Approach.* Hart (1995) maintains that (in contrast with the principal-agent approach) the property rights approach tries to address head-on the question of why there are less haggling and hold-up problems in a merged firm than between two independent, i.e., non-integrated firms. Why does ownership of physical or non-human assets matter? The answer, Hart (1995) submits, is that ownership is a source of control rights when contracts are incomplete.

Given that a contract will not specify all aspects of resource usage in every contingency, who has the property rights to decide about missing usages? According to the property rights approach, it is the owner of the resource in question who has these property rights. That is, the

---

<sup>8</sup> The modern property rights approach, discussed in Hart (1995), builds upon Grossman and Hart (1986), Hart (1989), and Hart and Moore (1990). Byrnjolfsson (1994) provides an insightful extension and application of the GHM model to information assets and information technology. Both extensions and critiques of the G-H-M model are many (see e.g., Holmstrom and Roberts, 1998; and Williamson, 2000). For the relevance of the property rights approach to strategic management see, for example, Liebeskind (1996) and Mahoney (1992a). Mahoney (1992a) notes an isomorphism between the Coase (1960) theorem that in the absence of transaction costs, liability rules do not matter for achieving economic efficiency and the idea expressed in the previous chapter that in the absence of transaction costs, organizational form (governance structure) does not matter for achieving economic efficiency. Of course, the main point of Coase (1937) is that in a world of positive transaction costs, organizational form choice does impact economic efficiency,

owner of a resource has *residual control rights* over the resource: the property rights to decide all usages of the resource in any way not inconsistent with a prior contract, custom, or law. In fact, possession of residual control rights is taken to be the definition of ownership in the modern property rights approach.

Hart (1995) concludes that the economic benefit of integration is that the acquiring firm's economic incentive to make relationship-specific investments increases since, given that the firm has more residual control rights, the firm will receive a greater fraction of the *ex post* surplus created by these relationship-specific investments. One implication of the property rights theory is that, *ceteris paribus*, a party is more likely to own a resource if he or she has an important (sunk cost) investment decision.

Another strategic implication of the property rights theory is that highly complementary assets should be under common ownership. For example, Joskow (1985) has investigated the ownership arrangements governing electricity-generating plants that site next to coal mines. Such relationship-specific assets are highly complementary, and Joskow (1985) finds a high incidence of common ownership. Stuckey (1983) has investigated the case of aluminum refineries that site next to bauxite mines. In this business situation, the degree of complementarity is arguably even greater, since, not only are the two entities located next to each other, but also the refinery installs equipment that is specific to the particular bauxite mine. Stuckey (1983) finds that vertical integration occurs in essentially every case. I submit that students studying the economics of organization who provide further case studies along the lines

---

and the main point of Coase (1960) is that in a world of positive transaction costs, initial property rights assignments do impact economic efficiency (as well as income distribution).

of Joskow (1985) and Stuckey (1983) that empirically test this modern property rights perspective would enrich the organizational economics research literature.

The Role of Non-human Assets and the Nature of Authority in Property Rights Theory.

The crucial economic features of the property rights approach are that contracts are incomplete, and that there are some significant non-human assets in the economic relationship. So far, we have focused on why contractual incompleteness is important to the modern property rights approach. We now consider why (at least some) non-human assets are an essential economic feature of a property rights theory of the firm. These non-human assets might include tangible assets such as machines, inventories, or buildings, or intangible assets such as patents, brand names, or the firm's reputation.

To understand better the role of non-human assets, consider a situation where firm 1 acquires firm 2, which consists entirely of human capital. What is to stop firm 2's workers from quitting? In the absence of any physical assets – e.g., buildings – firm 2's workers would not even have to relocate physically. For example, if the workers are linked by telephone or computer terminal (assets which the workers own themselves), workers could announce that they have become a new firm.

*For firm 1's acquisition of firm 2 to make any economic sense, there must be some source of firm 2's economic value over and above the workers' human capital.* This source of economic value may consist of: (a) a place to meet; (b) the firm's reputation, (c) a distribution network (assets that might be relevant to newspapers, journals or publishing houses); (d) the firm's files containing important information about its operations or its customers (assets that might be relevant for insurance companies or law firms); or (e) a contract that prohibits firm 2's workers from working for competitors or from taking existing clients with them when they quit (such a

contract may be relevant for accounting firms, public relations firms, advertising agencies, or R&D labs, as well as law firms). Thus, a firm's non-human assets represent the glue that keeps the firm together.

Hart (1995) notes that it is important to emphasize that there is no inconsistency between defining a firm in terms of non-human assets and recognizing that a large part of a firm's economic value derives from human capital. Suppose firm 2 consists of non-human asset  $a_2$  and one worker  $W_2$ . Assume that  $W_2$  can make \$300,000 a year using  $a_2$  and only \$200,000 in its absence, and suppose that  $W_2$  is the only person who knows how to operate  $a_2$  and that the scrap value of  $a_2$  is zero. Then, under the assumption of Nash bargaining, asset  $a_2$  is worth \$50,000 to an acquirer since the acquirer will be able to obtain 50% of  $W_2$ 's incremental \$100,000 by threatening to deny  $W_2$  access to the asset. That is, the economic value of the firm to an acquirer is significant even though the value of  $a_2$  in its next-best use (its scrap value) is zero.

Hart (1995) argues that the concept of non-human assets is also helpful for clarifying the concept of authority. Coase (1937), Simon (1947) and Williamson (1975) have argued that a distinguishing feature of the employer-employee relationship is that an employer can tell an employee what to do, whereas one independent contractor must explicitly compensate another independent contractor to do what he or she wants. However, as Alchian and Demsetz (1972) point out, it is unclear what is the source of an employer's authority over an employee. It is the case that an employer can tell an employee what to do, but it is also the case that one independent contractor can tell another independent contractor what to do. The pragmatically interesting question is why the employee acts accordingly, whereas the independent contractor (perhaps) does not pay attention.

When non-human assets are present there is a pragmatic difference between the employer-employee situation and the independent contractor situation. In the employer-employee case, if the employment relationship breaks down, the employer walks away with economically relevant non-human assets, whereas in the independent contractor case, each independent contractor walks away with non-human assets. This pragmatic difference gives the employer leverage. Put compactly: *Control over economically relevant non-human resources leads to control over human resources*. This argument connects the behavioral theory of the firm (March and Simon, 1958; Simon, 1947), transaction costs theory (Coase, 1937; Williamson, 1975), and modern property rights theory.

We next consider an application: the vertical merger of Fisher Body and General Motors (Klein, Crawford and Alchian, 1978) in light of modern property rights theory. Then, we begin the next chapter, which covers agency theory. We first consider the classic work by Berle and Means (1932) concerning the potential agency problem due to the separation of ownership and control. We then develop the basic foundations for the mathematical principal-agent model.

---

**Application: The Vertical Merger of Fisher Body and General Motors**

Source: Klein, B., V. Crawford and A. Alchian (1978). "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics*, (October): 297-326.

Originally, automobiles were constructed of open wooden bodies. By about 1919, however, closed metal bodies were being manufactured using giant presses to stamp the body parts. Making closed bodies required stamping dies that were in large measure specific to the particular requirements of the model to be produced. In the early period of the automobile industry, the producers of the dies were independent of the automobile manufacturers themselves. Soon after the shift toward closed bodies, which entailed a large specific investment on the part of the die manufacturers, long-term contracts appeared.

Because Fisher Auto Body had to develop specialized production devices that could only be used for General Motors (GM) cars, Fisher Body was reluctant to sign a short-term contract because at renegotiation time, Fisher Body would be at the mercy of General Motors. On the other hand, GM was reluctant to depend on so heavily on one supplier, fearing that, with a short-term contract, at renegotiation time, GM would be at the mercy of Fisher Body. Because each party feared a short-term contract would leave it at the mercy of the other firm, Fisher Body and General Motors signed a long-term contract for ten-years, according to which GM agreed to buy virtually all of its closed bodies from Fisher Body. This clearly protected Fisher Body from being held up by GM. But now opportunities have been created for Fisher Body to take advantage of GM. At what price would GM buy? Suppose demand conditions change greatly and GM wants to renegotiate the contract? How would quality be assured? Contract

negotiations became increasingly complex, until by 1926, the two firms merged, as a final attempt to mitigate bargaining difficulties, thereby replacing the transaction costs in the marketplace with internal organization. Vertical financial ownership replaced long-term contracting, which allowed the parties to adjust in an adaptive, sequential manner.

An important aspect of this case, based on the Grossman and Hart (1986) property rights theory of ownership is that much of the asset specificity came from investment in relationship-specific know-how by the Fisher Body workers, which would have made it difficult for General Motors to find another supplier if Fisher Body had tried to engage in holdup. Thus, vertical integration via financial ownership is persuasively explained in these property rights/ transaction costs terms.

---

## Chapter 4 Agency Theory

Agency theory is an influential approach to the study of corporate governance in strategic management (Eisenhardt, 1989; Kosnik, 1987; Oviatt, 1988; Rediker and Seth, 1995). While Berle and Means (1932) tended to be pessimistic about the economic effects of the separation of ownership and control, modern agency theorists (e.g., Fama, 1980; Fama and Jensen, 1983a; Jensen and Meckling, 1976) shift the emphasis from capital market failures to capital market efficiencies (Jaffe and Mahoney, 1999). Thus, modern agency theory is quite distinct, in this regard, from transaction costs theory.

Modern agency theorists tend to be (overly) optimistic that various governance mechanisms (e.g., the market for corporate control, the market for managers, and so on) have solved agency problems. In fact, some agency theorists (e.g., Fama and Jensen, 1983b) make the strong suggestion that these governance mechanisms have reached a high degree of refinement -- on which account there is not now, if indeed there ever has been, an organization control problem with which scholars and others are legitimately concerned.

I suggest that students, who are studying the economics of organizations, should have grave doubts about the overly optimistic (Chicago School) view that agency problems are solved. This view, in my judgment leads down a blind alley. This suggestion parallels the arguments in the previous chapter to be cautious in accepting the overly optimistic (Chicago School) view that property rights typically evolve toward economic efficiency. My own reading of the research literature is that the strong ideological fervor of Chicago School economists in espousing the virtues of the market has lead these economists more to wishful thinking than to coming to terms with agency (and property rights) problems that are well documented in the world of experience.



In addition to the descriptive agency theory of Jensen and Meckling (1976), and Fama and Jensen (1983a), there is also a more formal branch of agency theory where mathematical principal-agent models (e.g., Holmstrom, 1979) are developed. For the purposes of this research book, we focus more on the principal-agent model than on the descriptive theory. The Appendix to this chapter, however, covers the classic paper by Jensen and Meckling (1976).

The mathematical principal-agent problem in its moral hazard form stems from a basic conflict between insurance and economic incentives. On the one hand, the theory of optimal insurance demonstrates that the optimal division of a pie of a random size (e.g., the profit) between a risk-neutral party (the shareholders) and a risk-averse one (e.g., the manager) has the risk-neutral party bear all the risk, if economic incentives are left aside (see Tirole, 1988).

On the other hand, if the risk-averse party (the manager) is given an income that does not depend on effort, then the risk-averse party (the manager) has no economic incentive to exert effort. Thus, the goal of providing full insurance, conflicts with the goal of providing economic incentives. Specifically, for a risk-averse agent the insurance objective conflicts with the residual claimant status. Indeed, the trade-off between providing insurance and providing economic incentives generally leaves the contractual parties with both sub-optimal insurance and sub-optimal profits (a second-best contract).

In summary, the basic principal-agent model discussed in this chapter shows that: Efforts, if they are not observed, must be induced through economic incentives. The manager's compensation must grow with the realized profit. Because such economic incentive structures destroy insurance, the expected compensation is higher under non-observability. This fact, in turn, may make the principals (the shareholders) not wish to induce effort. Thus, the principals (the shareholders) may tolerate slacking.

Before focusing on this mathematical principal-agent model, we begin with a classic statement of the agency problem by Berle and Means (1932). We then consider the mathematical principal-agent model in Pratt and Zeckhauser (1985), Arrow (1985) and Levinthal (1988). Finally, we consider positive (i.e., descriptive) agency theory. Since the Chicago School optimistic view is mentioned throughout the agency theory chapter, the classic paper by Jensen and Meckling (1976) is included in the Appendix.

Berle and Means (1932) argue that the separation of ownership and control exists in varying degrees and that the separation of ownership and control becomes almost complete when not even a substantial minority shareholder exists. Berle and Means (1932) note, for example, that in the American Telephone and Telegraph Company, the largest shareholder owned less than one percent of the company's stock. Berle and Means (1932) raise concerns about the divergence of economic interests between shareholders and managers.

Pratt and Zeckhauser (1985) maintain that the issue of the separation of ownership and control is but a subset of a host of economic problems that can be classified as "principal-agent problems." Indeed, the theme of Pratt and Zeckhauser (1985) is that businesses, workers, and consumers regularly struggle to deal with agency problems. Overall, Pratt and Zeckhauser (1985) take a more optimistic view than Berle and Means (1932) concerning the severity of the agency problem, although Pratt and Zeckhauser (1985) do acknowledge the possibility that with creativity in the governance area, we could do better.

Similar to Pratt and Zeckhauser's (1985) arguments, Arrow (1985) notes that the agency relationship is a pervasive fact of economic life and that the principal-agent relationship has significant scope and economic magnitude. Arrow (1985) makes the useful distinction between

hidden action models (e.g., moral hazard models) and hidden information models (e.g., adverse selection models). Arrow (1985) also considers the role of multiple principals, the role of multiple agents, the role of monitoring, and the role of repeated relations in attenuating the agency problem. Finally, Arrow (1985) considers the extent to which the principal-agent relationships in actuality differ from that developed in the models. Most importantly, the theory tends to predict very complex fee functions to align economic interests between principals and agents. In fact, we do not observe such complex relations in reality. Although not mentioned by Arrow (1985), one obvious candidate to explain this divergence is that the mathematical-principal agent model assumes away the problem of bounded rationality.

Levinthal (1988) characterizes and critiques the research on agency models of organizations in order to broaden the set of readers of such models and to stimulate the production of new research. In many ways, this research paper highlights the ways in which the strategic management field can contribute to the evolving science of organization. This research paper uses specific mathematical functional forms that highlight the implications of these models. Moreover, Levinthal (1988) skillfully uses ideas from Simon (1947), March and Simon (1958) and Cyert and March (1963) to critique agency models from the perspective of the behavioral theory of the firm. A question left open to current students engaged in the evolving science of organization is whether agency theory and the behavioral theory of the firm will proceed in a dialectical manner with each posing questions and problems for which the other perspective must respond, or will a more synthetic approach emerge?

Finally, we conclude with the Chicago School view espoused by Jensen and Meckling (1976). Jensen and Meckling (1976) notes that if a manager has no debt or equity holdings in a firm, then the manager will bear none of the costs directly. In this situation, managers clearly

face an incentive to make decisions that are contrary to the best, wealth-maximizing interests of outside investors. Of course, outside investors may engage in a wide variety of activities to monitor the actions and decisions of managers. To go further, managers can post economic bonds whereby they will be penalized economically should they make decisions that violate the interests of outside investors. These continuing conflicts between managers and investors, together with the monitoring and bonding mechanisms designed to reduce these conflicts are costly.

Jensen and Meckling (1976) refer to the sum of these costs as *agency costs*. Total agency costs consist of (1) monitoring expenditures made by the principal to regulate and monitor the behavior of the agent, (2) bonding expenditures made by the agent to reassure principals, and (3) residual agency costs, or costs due to unresolved conflicts of interest between agents and principals --- this third category being a very expansive category. Jensen and Meckling (1976) argue that in a world of efficient capital markets, managers in an attempt to raise funds from outside sources bear all the economic burden of these agency costs. Jensen and Meckling (1976) use these conclusions to investigate the possibility of an optimal capital structure for a firm. Since managers bear all the agency costs in their attempt to raise capital, these managers have a strong economic incentive to choose the capital structure that minimizes total agency costs.

Thus, Jensen and Meckling (1976) argue that although the separation of ownership and control lessens profitability incentives, that incentive issue is fully anticipated at the time the separation of ownership from control occurs, and is therefore fully reflected in the price of new shares. The future therefore holds no surprises; all of the relevant contracting action is packed into ex ante incentive alignments.

Students should note that Jensen and Meckling (1976) examine the consequences of diluting a one hundred percent equity position in an entrepreneurial firm. However, their real theoretical interest is the diffusely owned modern corporation, but the basis for moving from one scenario to the other scenario is not described. Jensen and Meckling (1976: 356) expressly acknowledge this condition: “One of the most serious limitations of this analysis is that as it stands we have not worked out in this paper its application to the every large modern corporation whose managers own little or no equity. We believe our approach can be applied to this case but ... [these issues] remain to be worked out in detail.” As a general criticism, Williamson (1996: 188) commenting on Jensen and Meckling (1976) notes that: “the logic that connects tractable micro[economic] models and the composite uses to which they are put is often asserted but is rarely fully worked out.” Clearly, challenges to contemporary students are posed.

Berle, Adolf A. and Gardiner C. Means (1932). The Modern Corporation. New York, NY: Macmillan.

Berle and Means (1932) submit that corporations have ceased to be merely legal devices through which the private business transactions of individuals may be carried on. Berle and Means (1932) argue that the corporation has, in fact, become both a method of property tenure, and a means of organizing economic life.

For Berle and Means (1932) the direction of industry by persons other than those who have ventured their own wealth has raised the question of the motive force behind such direction, and the effective distribution of the economic returns from business enterprise. The private corporation has given way to an essentially different form, the quasi-public corporation; a corporation in which a large measure of the *separation of ownership and control* has taken place through the multiplication of owners.

Berle and Means (1932) argue that the separation of ownership and control produces a condition where the interests of owner(s) and managers may, and often do, diverge and where many of the checks that formerly operated to limit the use of such discretionary managerial power disappear. Physical control over the instruments of production has been surrendered in ever growing degree to centralized groups who manage the property in bulk, supposedly, but by no means necessarily, for the economic benefit of the security holders.

Berle and Means (1932) maintain that it has been assumed that, if individuals are protected in the rights both to use their property as they see fit and to receive the full fruits of its use, their desire for personal gains, for profits, can be relied upon as an effective economic

incentive to their efficient use of any industrial property they may possess. In the quasi-public corporation, such an assumption no longer holds.

Berle and Means take the position that where the separation of ownership and control is substantial, “control may be held by the directors or titular managers who can employ the proxy machinery to become a self-perpetuating body, even though as a group they own but a small fraction of the stock outstanding” (1932: 5). Berle and Means (1932) argue that ownership of wealth without appreciable control, and control of wealth without appreciable ownership, appear to be the logical outcome of corporate development. Berle and Means (1932: 121) ask: “Have we any justification for assuming that those in control of the modern corporation will also choose to operate it in the interests of the owners? The answer to this question will depend on the degree to which the self-interest of those in control may run parallel to the interests of ownership and, insofar as they differ, on the checks on the use of power which may be established by political, economic or social conditions”

Berle and Means (1932) in their empirical study found that 88 of the 200 largest American non-financial corporations to be “management controlled” in 1929 because no individual, family, corporation, or group of business associates owned more than 20 percent share of all outstanding voting stock, and because evidence of control by a smaller ownership group was lacking. Berle and Means (1932) judged only 22 of the corporations to be privately owned or controlled by a group of stockholders with a majority interest.

Berle and Means (1932) note that it is traditional that a corporation should be run for the economic benefit of its owners, the stockholders, and that to the stockholders should go any profits that are distributed. Berle and Means (1932) hold, however, that a controlling group may hold the power to divert profits into their own pockets. Berle and Means (1932: 333) submit

that: “There is no longer any certainty that a corporation will in fact be run primarily in the interests of the stockholders.”

*Berle and Means (1932) conclude that the rise of the modern corporation has brought a concentration of economic power that can compete on equal terms with the modern State — economic power versus political power.* The State seeks, in some respects, to regulate the corporation, while the corporation, steadily becoming more powerful, makes every effort to avoid (or shape) such regulation. Where its interests are concerned, the modern corporation even attempts to dominate the State. The future may see the economic organization, typified by the corporation, not only on an equal plane with the State, but possibly even superseding the State as the dominant form of social organization. The law of the corporation, accordingly, might well be considered as a potential constitutional law for the new economic State, while business practice is increasingly assuming the aspect of economic statesmanship.

We next consider an overview of the principal-agent research literature by Pratt and Zeckhauser (1985). This book provides us with why agency theory is relevant to contemporary management practice. Pratt and Zeckhauser (1985) emphasize that the real world has been inventive in developing subtle mechanisms, such as career expectations and product reputations, to overcome the difficulties posed by informational asymmetries. Thus, mathematical model builders need to learn from the world of business practice, and the insights provided by model builders can reciprocate providing insights for managerial practice.



Pratt, John W., and Richard J. Zeckhauser (1985). "Principals and Agents: An Overview" (pp. 1-35). In Principals and Agents: The Structure of Business, edited by John W. Pratt and Richard J. Zeckhauser. Boston, MA: Harvard Business School Press.

The Agency Relationship. Pratt and Zeckhauser (1985) note that a predominant concern for an economy, discussed since the time of Adam Smith (1776), is to assure that production is conducted in the most efficient manner, taking advantage of the economic benefits of specialization, while appropriately conserving scarce resources. But even if we could figure out, or were willing to let the market figure out, the most efficient way to produce goods and services, there would be the problem of ensuring that each individual performs his or her agreed-upon task. Full information rarely is freely available to all contractual parties, and so the economic problem becomes how to structure an agreement that will induce agents to serve the principal's interest even when the principal does not observe the agents' actions and information.

Whenever one individual depends on the action of another, an *agency relationship* arises. The individual taking the action is called the *agent*. The affected party is the *principal*. In common parlance, the doctor is the agent, and the patient is the principal. The corporate executive is the principal, and the subordinates are the agents. The corporate executive, in turn, is an agent for the shareholders.

Challenges in the agency relationship arise whenever the principal cannot perfectly and costlessly monitor the agent's action and information. The problems of inducement and enforcement then come to the fore. Given information asymmetries -- agents typically know more about their tasks than their principals do -- we cannot expect any business enterprise or business institution to function as well as it would if all information were costlessly shared or if the economic incentives of principals and agents could be costlessly aligned. This shortfall is

sometimes called the *agency loss* or *agency costs*. The economic challenge in structuring an agency relationship is to minimize agency costs. In economic parlance, since the first-best outcome could be achieved only in the unrealistic world of costless information flow, our goal must be to do the best we can, to achieve what is sometimes called the *second-best solution*.

Pratt and Zeckhauser (1995) note that businesses, workers, consumers, and indeed all participants in society at large regularly struggle to deal with the intractable problems that arise in agency relationships, that organizational forms evolve to deal with such agency costs, and on average, these organizational forms perform reasonably well. The building blocks of agency theory are information and economic incentives.

Information. Pratt and Zeckhauser (1985) note that in most social and business relationships, the contractual parties have different information available to them. Relationships vary in the degree of informational asymmetry they involve. At one extreme we have the perfect-market transaction, with standardized products and all information fully shared. At the other end of the continuum are situations in which the agent has full discretion and is not observed at all by the principal.

Monitoring. Pratt and Zeckhauser (1985) provide several agency theory predictions concerning monitoring:

- We tend to get less monitoring, or monitoring of poorer quality, when monitoring is expensive and/or substitutes for monitoring are cheap;
- The agency loss is more severe when the economic interests or economic values of the principal and agents diverge substantially, and information monitoring is costly;
- In a range of real-world situations, more limited monitoring -- say of an indicator of output -- is relatively or fully successful;

- A large stock of economic value that could be lost through bad behavior, such as reputation or assets subject to legal suit, is a strong economic incentive for good behavior;
- Long-term relationships, among other benefits, develop the stocks of economic value needed for "enforcement," and make limited monitoring more effective;
- The economic benefits of any reductions in agency loss will be shared by principal and agent in most market situations; and
- The principal and agent have a *common economic interest* in defining a monitoring-and-incentive structure that produces economic outcomes as close as possible to the economic outcome that would be produced if information monitoring were costless.

Incentives. Pratt and Zeckhauser (1985) note that each person performs his or her task in Adam Smith's (1776) pin factory. The specialization of labor that has created the possibilities for productive modern industrial societies has also created the need for organizations larger than the crafts shop or family. The new challenge becomes how to motivate the participants within the organizations that make up society so that these participants will be as productive as they would be if they were the owners.

Pratt and Zeckhauser (1985) argue that the simple prescription of neoclassical microeconomic theory is difficult to apply in business practice. It is difficult to measure an individual's marginal product. If Company A's sales go up, is it because of the perseverance of the salesperson, the workers' greater attention to quality, or blind luck? The problem of measuring marginal product has become so intractable that the vast majority of Americans are paid based on salaries. If work input is difficult to measure, the employee may have an economic incentive to shirk. Some indicator of inputs, such as hours on the job, is likely to be used in evaluating the employee instead of a measure of actual inputs, which would include the diligence

and quality of those hours. Contemporary wage compensation systems, in which compensation is not closely tied to productivity, pose a sizeable agency problem for the motivation of workers.

Pratt and Zeckhauser (1985) maintain that the corporation and the modern economy have found ways to respond to this separation of rewards and productivity. Part of the solution is monitoring by supervisors. Promotions, for example, offer an individual higher future wages for the indefinite future. And dismissal typically imposes a major cost on the worker. If motivation is particularly important and monitoring is costly, wages may be set above the amounts workers could earn elsewhere; the small chance of termination will then be an influential economic incentive (Klein and Leffler, 1981).

Pratt and Zeckhauser (1985) note that unlike labor, capital need not be motivated. But informational asymmetries still can create agency problems for capital. The quality of a machine can be overstated; consider the used car or the new computer. Machines that can be abused or overused, but whose condition is difficult to monitor, are sold rather than leased. And a production process is likely to be vertically integrated if there are severe quality uncertainties at various supply-chain stages.

Modern accounting attempts to keep track of capital and the profits such capital produces. However, accounting techniques cannot accurately assess many contributors to a firm's long-term profitability, such as reputation for quality, condition of equipment, or research accomplishments. An agency loss may be the consequence, with managements pursuing measured outputs, such as reported profits, at the expense of those accomplishments that are difficult to tally. Some critics of modern business enterprises have identified this economic distortion as a major factor contributing to lagging American productivity.

Pratt and Zeckhauser (1985) emphasize that the types of informational asymmetries that lead to agency loss make it difficult to reward labor or capital with the economic value of its product. Having the factors of production in a single business entity such as a corporation is advantageous. Long-term residence there allows for more creative reward structures, such as internal promotion opportunities. Long-term relationships also encourage individuals to work toward a common purpose.

The Design of Agency Structures. Pratt and Zeckhauser (1985) note that in considering the agency problem, it is perhaps natural to focus on the question of how the principal can reap the greatest advantage through economic incentives that influence agents' behavior, yet reward the agents enough so that these agents will not quit. Two central assumptions are implicit in this formulation: first, that the principal is in a position to design the monitoring and incentive mechanism; and second, that all the economic benefits from improvements in performance go to the principal. Pratt and Zeckhauser (1985) argue that agency situations often satisfy neither of these assumptions, and that the most important issues do not depend on them. In fact, the agent and principal are merely two (or more) individuals (or organizations) in some sort of explicit or implicit contractual relationship. Oversight and reward mechanisms may be designed by either the agent (e.g., most lawyers define the terms of their relationships with clients) or the principal (e.g., the store manager probably sets the employment conditions for retail clerks). There may be a joint negotiation, as we see with labor agreements under collective bargaining. Or some external party may set the terms, as the government does in structuring many of the rules by which corporations are governed in the United States.

*Pratt and Zeckhauser (1985) conclude that while agency theory has provided modest aid in improving business practice, the real world of business has been inventive in developing*

*subtle mechanisms such as career expectations and product reputations to overcome the difficulties posed by informational asymmetries.* Also, human environments can change quickly and there is no assurance that the institutions we currently observe are best. Therefore, sound conceptual thinking concerning institutions by students studying the economics of organization can be worthwhile for reducing agency loss in the future.<sup>9</sup>

We next consider Arrow's (1985) review of the principal-agent model. Arrow (1985) also concludes with a thoughtful evaluation of the agency model.

---

<sup>9</sup> For example, a number of mechanisms have evolved to attenuate the agency problem (e.g., the market for corporate control; the market for managers, use of outside board of directors, monitoring by institutional investors, concentrated ownership, and so on (see Mahoney, 1992a). For both economic and strategic management theoretical and empirical contributions to the agency theory literature see Eisenhardt (1985, 1989), Fama (1980), Fama and Jensen (1983a, 1983b), Holmstrom (1979), Jensen and Meckling (1976), Kosnik (1987), Lajili, Barry, Sonka and Mahoney (1997), Mahoney and Mahoney (1993), Mahoney, Sundaramurthy and Mahoney (1996, 1997), Rediker and Seth (1995), Seth (1990), Seth and Thomas (1994), Sundaramurthy and Lewis (2003); Sundaramurthy, Mahoney and Mahoney (1997), Walsh and Seward (1990) and Zenger (1994).

Arrow, Kenneth J. (1985). "The Economics of Agency" (pp.37-51). In Principals and Agents: The Structure of Business, edited by John W. Pratt and Richard J. Zeckhauser. Boston, MA: Harvard Business School Press.

Arrow (1985) observes that the agency relationship is a pervasive fact of economic life. The principal-agent relation is a phenomenon of significant scope and economic magnitude. A common element for early principal-agent models is the presence of two individuals. One (the agent) must choose an action from a number of alternatives. The action affects the economic welfare of both the agent and another person, the principal. The principal, at least in the simplest cases, has the additional function of prescribing payoff rules; that is, before the agent chooses the action, the principal determines a rule that specifies the fee to be paid to the agent as a function of the principal's observations of the results of the action.

The economic literature has focused primarily (but not exclusively) on the case in which: (1) The agent's action is not directly observable by the principal, and (2) the outcome is affected but not completely determined by the agent's action. Were it not for the second condition, the principal could infer the agent's action by observing the outcome. In more technical language, the outcome is a random variable whose distribution depends on the actions taken.

Arrow (1985) notes that more generally, a single principal may have many agents. Each agent takes an action, and the output of the system is a random function of all the actions. The principal cannot observe the actions themselves but may make some observations, for example, of the output. Again, the principal sets in advance a schedule stating the fees to be paid to the individual agents as a function of the observations made by the principal. The principal-agent theory is in the standard economic tradition. Both principal and agent are assumed to be making their decisions optimally in view of their constraints, and intended transactions are realized. As is

usual in economic theory, the theory functions both normatively and descriptively. The theory offers insights used in the constructions of contracts to guide and influence principal-agent relations in the real business world, and at the same time the theory represents an attempt to explain observed phenomena in the empirical economic world, particularly exchange relations that are not explained by more standard economic theory.

Arrow (1985) finds it useful to distinguish two types of agency problems, *hidden action* (e.g., moral hazard) and *hidden information* (e.g., adverse selection). A typical hidden action is the effort of the agent. Effort is a disutility to the agent, but the agent's effort has an economic value to the principal in the sense that such effort increases the likelihood of a favorable outcome (technically, the distribution of the outcome to a higher effort stochastically dominates that to a lower effort; that is, the probability of achieving an outcome that exceeds a given level is higher with higher effort). An example of hidden action is the relation between stockholders and management. The stockholders are principals, who cannot observe in detail whether the management, their agent, is giving high effort and is making appropriate decisions.

Fire insurance dulls economic incentives for caution and even creates economic incentives for arson; this economic problem is the origin of the term moral hazard. Health insurance creates similar economic problems because health insurance creates an economic incentive to use excessive medical care. Employment relationships typically are relationships in which effort, and ability acquired through training and self-improvement are difficult to observe. In one view, firms exist as a means of measuring effort (Alchian and Demsetz, 1972).

In hidden information problems, the agent has made some observation that the principal has not made. The agent uses (and should use) this observation in making decisions; however,



the principal cannot check whether the agent has used his or her information in the way that best serves the principal's interest. A case much studied from various points of view in the economic research literature is that of a decentralized socialist economy. The productive units may well have economic incentives not to reveal their full potentiality, because it will be easier to operate with less stringent requirements. The economic problem for the central planning unit (the principal) is how to tap the agent's information. A similar problem occurs in decentralization within a firm. This topic in the literature has acquired the name of *incentive compatibility* (Miller, 1992).

The problem of adverse selection was originally noted in insurance of several kinds. The population being insured is heterogeneous with respect to risk; in the case of life insurance, for example, some have a higher probability than others of dying young. In at least some cases, those who are insured have better knowledge of this probability than the insurance company, which is unable to differentiate. If the same premium is charged to everyone, then the high-risk individuals will purchase more insurance, and the low-risk individuals will purchase less insurance. This outcome leads to an inefficient allocation of risk bearing. Arrow (1985) notes that the hidden information principal-agent model becomes more complicated when multiple principals compete for agents.

The Hidden Action Model. Arrow (1985) provides a simple formulation of the hidden action model. The agent (for the moment, assume there is only one) chooses an action  $a$ . The result of this choice is an outcome  $x$ , which is a random variable whose distribution depends upon  $a$ . The principal has chosen beforehand a *fee function*  $s(x)$ , to be paid to the agent. For the simplest case, assume that the outcome  $x$  is income -- that is, a transferable and measurable quantity. Then the net receipts of the principal will be  $x - s(x)$ . Because the principal and agent

are both, in general, risk averse, each values whatever income he or she receives by a utility function with diminishing marginal utility. Let  $U$  be the utility function of the principal, and  $V$  that of the agent. Further let  $W(a)$  be the disutility the agent attaches to action  $a$ . It will be assumed separable from the utility of income; that is the marginal utility of income is independent of the action taken (the amount of effort). Note that the action is taken before the realization of the uncertainty and is therefore not uncertain to the agent, though the agent's action is unknown to the principal.

Since even for a given action, the outcome  $x$ , is uncertain, both principal and agent are motivated to maximize the expected value of their own utility. Given the principal's choice of fee function  $s(x)$ , the agent wishes to maximize the expected value of  $V[s(x)] - W(a)$ . In effect, therefore, the principal can predict the action taken for any given fee schedule. The choice of fee schedules is, however, restricted by competition for agents, who have alternative uses for their time. Hence, the principal must choose a fee schedule that offers the agent a utility at least equal to what the agent could achieve in other activities (i.e., there is a participation constraint).

*The principal-agent problem combines two inextricable elements: risk sharing and differential information.* Even if there were no economic problem of differential information, there would be some sharing of the outcome if both contractual parties are risk averse. Indeed, if the agent was risk neutral, the principal-agent problem would have a straightforward solution, the agent would bear all the risks, and then the differential information would not matter. That is, the principal would retain a fixed amount and would pay the remainder to the agent, who therefore would have no dilution of economic incentives. In the terminology used above, the fee function would equal the outcome less a fixed amount,  $s(x) = x - c$ , where the constant  $c$  is

determined by the participation constraint. Thus, a landlord renting land to a tenant farmer would simply charge a fixed rent independent of output, which in general depends on both the tenant's effort, which is unobservable to the landlord, and the vagaries of the weather.

However, this economic solution is not optimal if the agent is risk averse. Since all individuals are averse to sufficiently large risks, the simple solution of preserving economic incentives by assigning all risks to the agent fails as soon as the risks are large compared with the agent's wealth. For example, the president of a large company can hardly be held responsible for the company's income fluctuations.

In the general case of a risk-averse agent, the fee will be a function of the outcome, in order to provide economic incentives, but the risk will be shared. Generally, there is a trade-off between economic incentives and the efficiency of the system (considering both principal and agent). For a business application, consider the case of insurance with moral hazard. Some insurance will be written, but insurance will not be complete. In the terminology of the insurance industry, there will be *co-insurance*; that is, the insured will bear some of the economic losses against which the insurance is written. Co-insurance is customary in health insurance policies where the insured has considerable control over the amount of health expenditures.

Monitoring. Arrow (1985) notes that some principal-agent research literature has emphasized the possibility of monitoring. By this is meant the possibility that the principal has certain information in addition to the outcome. If this observation  $y$  is costless and conveys any information about the unobserved action  $a$  beyond that revealed by the outcome,  $x$  (technically, if  $x$  is not a sufficient statistic for the pair  $x, y$  with respect to action  $a$ ) then the principal can always improve the second-best contract by making the fee depend upon  $y$  as well as  $x$ .

Multiple Agents and Repeated Relations. Arrow (1985) maintains that new possibilities for economic incentives arise when there are many agents for a single principal, or alternatively, repeated relations between agent and principal. One can use the ordinal ranking of agents' outcomes as a basis for fees.

Repeated relations between a principal and agent provide new opportunities for economic incentives. Experience rating in insurance illustrates the situation; the premium rate charged today depends on past outcomes. In effect, the information on which the fee function is based is greatly enriched.

Evaluation of Agency Theory. Arrow (1985) submits that principal-agent theory gives a good reason for the existence of sharecrop contracts, but principal-agent theory is a very poor guide to their actual terms. Indeed, as John Stuart Mills pointed out long ago, the terms tend to be regulated by custom. Sharecrop contracts are remarkably uniform from farm to farm, and from region to region. Principal-agent theory, in contrast, suggests that the way the produce is divided between landlord and tenant would depend on the probability distribution of weather and other exogenous uncertainties, and on the relations between effort and output both of which certainly vary from one region to another. The relation between effort and output would vary over time as well. Similarly, the co-insurance provisions in health insurance policies are much simpler than could possibly be accounted for by principal-agent theory.

Arrow (1985) observes that in some cases where principal-agent theory seems clearly applicable, real-world business practice is very different from the model. In many respects, the physician-patient relation exemplifies the principal-agent relationship almost perfectly. The principal (the patient) is certainly unable to monitor the efforts of the agent (the physician). The

relation between effort and outcome is random, but presumably there is some connection. Yet, in practice, the physician's fee schedule is in no way related to outcome. In general, compensation of professionals shows only a few traces of the complex fee schedules implied by agency theory.

Why is this divergence between theory and practice so stark? Arrow (1985) argues that one basic problem is the costs of specifying complex relations. Second, superiors judge executives on criteria that could not have been stated in advance. Outcomes and even supplementary objective measures simply do not exhaust the information available on which to base rewards. A third limitation of the present model is the restricted reward or penalty system used. It is virtually always stated in terms of monetary payments. *Further extensions of the agency model are needed to capture some aspects of reality, for there is a whole world of rewards and penalties that take social rather than monetary forms.* Arrow (1985) submits that professional responsibility is clearly enforced in good measure by systems of ethics, internalized during the education process and enforced in some measure by formal punishments and more broadly by reputations. Ultimately, of course, these social systems have economic consequences, but they are not the immediate ones of current principal-agent models.

We conclude the agency chapter with an application by Wolfson (1985), followed by an appendix that contains Levinthal's (1988) instructive paper on principal-agent models and their limitations and Jensen and Meckling's (1976) classic paper on agency theory.

---

## **Application: Empirical Evidence of Incentive Problems and Their Mitigation**

Source: Wolfson, M. A. (1985). Empirical evidence of incentive problems and their mitigation in oil and gas shelter programs (pp. 101 – 125). In J.W. Pratt and R. J. Zeckhauser (Eds.), Principals and Agents.

Wolfson (1985) examines the oil and gas industry, a business context with high agency problems due to asymmetric information. Moreover, the United States tax code has encouraged investment in oil and gas drilling by limited partnerships --- a classic agency relationship. The limited partners (the principals) put up most of the capital, while the general partner (the agent) makes most of the decisions. The limited partners are typically far from the scene of the action, and most of them have no experience or expertise in the business.

There are severe conflicts of interests in this business context. The general partner frequently sells services to his limited partners. The general partner may also have related investments nearby and may benefit personally from the information gained from drilling. Moreover, the partnership typically imposes all of the immediately deductible costs on the limited partners (to capitalize on the tax advantages), whereas completion costs are shared between the limited and general partner, or may even be fully the responsibility of the general partner. Therefore, the general partner has too much economic incentive to drill and, for some revenue splits, too little economic incentive to complete.

The agency problems in drilling are well known to both operators and limited partners. Prospectuses detail these agency problems at length, and point out that monitoring by the limited partners is not really feasible. To reassure prospectus investors, general partners explain that they are most concerned about maintaining their long-term business reputation.

Wolfson (1985) tests a number of empirical assertions about the economic outcomes that, based on agency theory, we might expect in this business context. For example, the problem of distorted economic incentives is less severe in exploratory than developmental drilling. Hence, many limited partnerships should operate in the business context of exploratory drilling, and this economic outcome does hold empirically.

Wolfson (1985) also investigates the relationship between reputation and reduced agency loss. If an operator general partner can build a reputation for success, Wolfson (1985) shows empirically that the general partner can secure investment funds on more favorable economic terms. To build reputation, the general partner takes actions to benefit the limited partners at the expense of the general partner's own immediate profitability. Wolfson's (1985) data reveal that operator general partners with records of success for limited partners do receive more favorable economic terms from investors. Wolfson (1985) shows that even, or perhaps particularly, in the potentially opportunistic business environment of oil and gas drilling, business reputations can help overcome contracting difficulties and the resulting economic inefficiencies.

---

## Appendix:

Levinthal, Daniel (1988). "A Survey of Agency Models of Organizations." Journal of Economic Behavior and Organization, 9: 153-185.

Levinthal (1988) provides the insightful perspective that agency theory can be viewed as the neoclassical economic response to the questions raised many years earlier by March and Simon (1958) and Cyert and March (1963) regarding the behaviors of an organization of self-interested agents, with conflicting goals, in a world of incomplete information.

An agency relationship is said to exist between two (or more) parties when one contractual party, designated as the agent, acts on behalf of another contractual party, designated the principal. For example, in an attempt to bring management's interests in line with those of stockholders, top management is given complex compensation packages, consisting of salary, performance bonuses, stock ownership, and pension plans.

Agency theory views the economic problem of contract design as maximizing the economic payoff to the principal, taking several factors into account:

- The relationship between output and the economic incentive scheme offered;
- The allocation of risk associated with different compensation schemes; and
- The preferences of the principal and agents with respect to income and non-pecuniary outcomes.

### Agency Model

Consider an organization consisting of two utility maximizing individuals. One is the owner, the other a hired manager. The owner's economic problem is to design a compensation package that elicits an appropriate effort level from the manager. This economic problem is complicated by the fact that the owner only has limited information. The owner only observes the firm's revenue, which is a function of two factors that the owner cannot directly observe: (1)



the manager's effort, and (2) some exogenous variable such as consumer demand. Given the observed revenue level, the owner can only make probabilistic inferences about the manager's effort level. To make the example more concrete, suppose that Table 1 indicates the relationship between revenue, the manager's effort,  $a$ , and the exogenous state variable  $S_i$ .

**Table 1**

		State				Expected Cash Flow
		$S_1$	$S_2$	$S_3$	$S_4$	
Effort Level:						
$a$ :	<i>High</i>	50,000	50,000	50,000	25,000	43,750
	<i>Medium</i>	50,000	25,000	50,000	25,000	37,500
	<i>Low</i>	50,000	25,000	25,000	25,000	31,250

For instance if the manager exerts high effort (High) and state  $S_4$  occurs, then the cash flow is \$25,000. The expected cash flow from the agent's effort level is derived by assuming that the four states are equally likely to occur; that is the expected cash flow corresponding to a high level of effort is computed as  $0.75 (\$50,000) + 0.25 (\$25,000) = \$43,750$ . The agent's effort is assumed to be productive, in the sense that the expected cash flow increases as the agent's effort level rises from Low to Medium to High.

More generally, we can think of the observed outcome (revenue in the above example) as a random variable whose distribution depends on the action (effort) chosen by the agent. Let  $f(x | a)$  represent a probability distribution of  $x$ , conditional on the agent's action. Table 1 presents an example of a particular distribution of  $f(x | a)$ . The usefulness of the agent's efforts is modeled by the assumption that increased effort shifts the distribution of outcomes to the right in the sense of first-order stochastic dominance. This characteristic of the model is illustrated in Table 1 where for any state  $S_i$ , increased effort weakly increases revenue (i.e., the revenue either increases or stays the same).

Since the owner cannot directly enforce a particular effort level, the owner must influence the agent's self-interest by the choice of compensation scheme. This choice of economic incentive structure is also constrained by the fact that the owner cannot force the manager to work for the firm. As a result, the manager's alternative employment opportunities must be considered --- a participation constraint, to use Arrow's (1985) term.

In order to characterize the optimal compensation scheme for this example, we must specify the manager's preferences. It is assumed that the manager has preferences over income and effort represented by a utility function  $W(z, a) = U(z) - V(a)$ , where  $z$  is the level of income and  $a$  is the level of effort. It is assumed that the utility function is separable into the utility for money  $U(z)$  and the disutility of effort  $V(a)$ .

In particular,  $U(z) = \sqrt{z}$  and

$V(a) =$	$V(\text{high}) =$	40
	$V(\text{medium}) =$	20
	$V(\text{low}) =$	05

The function  $U(z)$  is one in which the agent's utility increases with  $z$  but at a diminishing rate (i.e., this is a concave function) and implies that the agent is risk averse. Secondly, for a given income level, the agent's utility is decreasing in the agent's effort level. More generally in agency analyses, it is assumed that  $U(z)$  is concave and  $V(a)$  is convex with respect to the continuous effort variable. The agent is risk averse, and the agent's disutility of effort increases at an increasing rate. In addition, we need to state the agent's opportunity cost of working on behalf of the principal. In our example, we assume that the agent's best alternative yields a utility measure of 120 units.

Finally, to complete the model, we must specify the owner's preferences. It is assumed that the owner derives utility only from the net earnings, i.e., the cash flow specified in Table 1 less the agent's compensation. The owner's utility function is represented by  $G(\cdot)$ ; in the example, we assume that the principal is risk-neutral and  $G(x) = x$ .

Suppose that the owner wants to elicit an effort of *high* from the manager. The owner faces two constraints. First, the contract must yield the manager an expected utility of at least 120 units if the manager is to work in the owner's behalf. Secondly, it must be in the manager's own interest to choose an effort level of high over medium over low. The only device under the owner's control is the compensation scheme that the owner offers the manager. The agent's share of compensation is not a fixed payment, but a schedule of payments expressing the agent's reward as a function of the revenue of the firm. The agent's compensation is represented by the sharing rule  $z(x)$ . Let  $Z_{50}$  be the payment to the manager if the outcome  $x = \$50,000$  is observed, and  $Z_{25}$  the payment if the outcome  $x = \$25,000$  is observed. The expected utility of a manager who sets  $a = \text{high}$  is the following:

$$E[U(\text{high})] = 0.75(\sqrt{z_{50}}) + .25(\sqrt{z_{25}}) - 40$$

$$E[U(\text{medium})] = 0.50(\sqrt{z_{50}}) + .50(\sqrt{z_{25}}) - 20$$

$$E[U(\text{low})] = 0.25(\sqrt{z_{50}}) + .75(\sqrt{z_{25}}) - 5$$

If the manager is to find it to his or her interest to choose  $a = \text{"high effort"}$  over the other alternative effort levels then the following relationships must be satisfied:

$$E[U(\text{high})] > E[U(\text{medium})]; \text{ and}$$

$$E[U(\text{high})] > E[U(\text{low})]$$

In addition, if the manager is to accept employment with the owner then

$$E [U (\text{high})] \geq 120$$

We can write the owner's decision problem as the following constrained optimization problem:

$$\text{Min } 0.75 (Z_{50}) + 0.25 (Z_{25})$$

$$\text{st. } 0.75 (\sqrt{z_{50}}) + .25 (\sqrt{z_{25}}) - 40 \geq 120$$

$$0.75 (\sqrt{z_{50}}) + .25 (\sqrt{z_{25}}) - 40 \geq .50 (\sqrt{z_{50}}) + .50 (\sqrt{z_{25}}) - 20$$

$$0.75 (\sqrt{z_{50}}) + .25 (\sqrt{z_{25}}) - 40 \geq .25 (\sqrt{z_{50}}) + .75 (\sqrt{z_{25}}) - 5$$

The solution to this optimization problem is  $Z_{50} = \$32,400$  and  $Z_{25} = \$10,000$ .

However, the owner's calculations are not over. This economic incentive scheme is the solution to the economic problem of how an owner can elicit an effort level of high with least cost, but high effort may not be the optimal choice from the owner's perspective. While it is true that high effort generates the highest expected revenue, the cost of eliciting this effort level may be such that the additional revenue is not worth the additional cost. In fact, high effort is not optimal in this mathematical principal-agent model example.

The economic incentive scheme that at minimum expected cost motivates the agent to choose medium effort is to set  $Z_{50} = \$28,900$  and  $Z_{25} = \$12,100$ . This economic incentive scheme yields the owner expected net earnings of  $0.5 (\$50,000 - \$28,900) + 0.5 (\$25,000 - \$12,100) = \$17,000$ .

In contrast, the expected net earnings from the cost minimizing scheme that elicits an effort of high is  $0.75 (\$50,000 - 32,400) + 0.25 (\$25,000 - 10,000) = \$16,950$ . The increase in

the expected compensation required to elicit this higher effort level dominates the increase in revenue.

Finally, if the principal wishes to elicit a low effort the optimal payment scheme is  $Z_{50} = \$15,625$  and  $Z_{25} = \$15,625$ . Since the agent's utility is decreasing in the agent's effort level, paying the agent a constant wage elicits the minimum effort level. In addition, the constant wage is the most efficient risk-sharing arrangement, since the manager is risk-averse and the owner is risk-neutral.

In order to elicit a higher effort level, the manager must bear some of the uncertainty regarding the state of nature. This trade-off between providing motivation and risk sharing is a general feature of agency problems. In order to provide an economic incentive for the manager to choose anything but the minimum effort level, the compensation schedule must deviate from the efficient risk-sharing contract (Holmstrom, 1979).

Levinthal (1988) notes that it is instructive to contrast the optimal contract under incomplete information with that under complete information. In this example, the fact that the owner is risk-neutral and the manager risk averse, suggests that in the first-best contract all the risk regarding the state of nature should be borne by the owner. An alternative arrangement is inefficient, since the manager is willing to take a salary reduction in order to insure against uncertainty in the manager's income, while the risk-neutral owner is indifferent to any uncertainty in income.

As a result, the optimal compensation scheme is calculated under the assumption of complete information. For instance, if the manager is to exert a high effort level, the manager must receive sufficient compensation to yield a net utility level of 120 units. Therefore, if  $z$  is the

wage, then it must satisfy the following relationship:  $\sqrt{z} - V(\text{high}) = 120$ . Since  $V(\text{high})$  is equal to 40 utility units, then  $Z = \$25,600$ . Table 2 indicates the required payment level for all three of the effort levels and the resulting net earnings of the owner.

**Table 2**

<b>Effort</b>	<b>Expected Revenue</b>	<b>Required Wage</b>	<b>Net Earnings</b>
<i>High</i>	\$43,750	\$25,600	\$18,150
<i>Medium</i>	\$37,500	\$19,600	\$17,900
<i>Low</i>	\$31,250	\$15,625	\$15,625

Examining the owner's net earning reveals that high effort is the preferred effort level, yielding the owner an expected net income of \$18,150 (with complete information and where the agent's effort is directly observable). Contrast this first-best contract under complete information with the optimal contract under incomplete information. There are two losses compared to the first-best solution: *first, the manager's effort level is reduced from the first-best level; second, there is inefficient risk-sharing, since the manager bears some of the risk associated with the exogenous state of nature.* The second-best compensation scheme provides an economic incentive for the agent to work, but this compensation scheme comes at the expense of the benefits of risk sharing. Agents are, to some extent, held accountable for events over which they have no control.

Levinthal (1988) notes that the fundamental insight of the characterization of the second-best contract relates to the imposition of risk on the agents. The first-best contract is analogous to a wage contract. In jobs where supervisors can directly observe labor input, a first-best wage contract is possible, whereas, in jobs in which the worker's input is not readily observable, the

optimal contract imposes risk on the worker. The latter case is clearly applicable to managerial and sales personnel. For these workers, compensation is frequently based on uncertain outcomes. In the case of managers, compensation is often based on divisional profits; similarly, for sales personnel, compensation may be based on the number of products sold. These outcomes are determined in part by the worker's actions, but are also significantly influenced by factors beyond the worker's control. However, given the non-observability of the worker's actions, compensation must be based on these partially random outcome measures if the worker is to have an economic incentive to choose any but the lowest level of effort.

Levinthal (1988) points out that the first-best contract can be achieved in a trivial manner if the agent is risk neutral. Risk neutrality of the agent implies that there is no economic welfare loss in having the agent absorb all the risk associated with the uncertain outcome. In this case, the optimal contract takes the form of the principal receiving a fixed payment and the agent receiving the residual outcome. Essentially, this amounts to the agent purchasing the firm from the principal.

Monitoring. Levinthal (1988) notes that there is a trade-off between imposing risk on the agent and carrying out costly monitoring activities. Thus, one can view inefficient risk sharing as a *substitute* for monitoring. For example, a sales commission is a substitute for measuring how hard and cleverly a salesperson works. Methodologically, this substitution effect has the strategic implication that the choice of monitoring mechanism and compensation scheme must be considered jointly.

Consider an information system that provides imperfect information as to the state of nature. The reduction in risk imposed on the agent allows the principal to reduce the expected compensation to the agent and still satisfy the requirement that the agent's expected wage is at

least as large as the agent's opportunity cost. Monitoring helps the principal distinguish whether the outcome results from the agent's action or the state of nature.

Levinthal (1988) points out that an agency theory setting is a non-cooperative game in which the actions are chosen by players in the game, rather than by nature as in a statistical decision problem. That is, agency theory is *game-theoretic*, rather than *decision-theoretic*.

Role of Time. Levinthal (1988) notes that the repetition of an agency relationship over time tends to improve its efficiency. Holmstrom suggests that: "when the [agency relationship] repeats itself over time, the effects of uncertainty tend to be reduced and dysfunctional behavior is more accurately revealed, thus alleviating the problem of moral hazard" (1979:90).

Multi-agent models and tournament contracts. Levinthal (1988) maintains that the risk imposed on an agent can be reduced by basing individual compensation on individual performance relative to that of other agents, who face similar states of nature. For example, in "tournaments," the reward is a function of the rank order of performance relative to other agents.

Conclusions. Levinthal (1988) concludes that the economic incentive conflict between stockholders and management does not appear to stem primarily from effort aversion on the part of management. *It is not the lack of industriousness of top management that is typically at issue, but the qualitative nature of the decisions that managers make.* For instance, it is frequently claimed that management's risk aversion leads to more conservative behavior than the maximization of shareholder wealth would imply. Models based on effort aversion may be appropriate in some contexts, but the emphasis placed on them in the principal-agent research literature is excessive. The extent of the relevance of this research literature to the economics of



organization depends on the connections that the agency theory literature makes to empirical work (see e.g., Eisenhardt, 1988).

We conclude the agency theory chapter with Jensen and Meckling (1976). This paper provides a positive (i.e., descriptive) theory of agency issues. Jensen (1983) refers to the “positive theory of agency.” Here, Jensen argues that: “capital intensity, degree of specialization of assets, information costs, capital markets, and internal and external labor markets are examples of factors in the contracting environment that interact with the costs of various monitoring and bonding practices to determine contractual forms” (1983: 334-335). This positive agency branch repeatedly asserts that natural selection processes are reliably efficacious (Fama, 1980; Fama and Jensen, 1983a, 1983b; Jensen, 1983).

My own reading is that the positive agency theory (e.g., Fama, 1980; Jensen and Meckling, 1976; Jensen, 1983) provides a far too optimistic view concerning the attenuation of conflicts of interests between agents and principals. Be that as it may, the Chicago School view is given its voice in the final section of this agency theory chapter.

## Appendix:

Jensen, Michael and William Meckling (1976). Theory of the firm: Managerial behavior, agency costs, and capital structure. Journal of Financial Economics, 3 (October): 305-360.

Jensen and Meckling (1976) integrate elements from agency, the theory of property rights and the theory of finance to develop a theory of the ownership structure of the firm. Jensen and Meckling (1976) define an agency relationship as a contract under which one or more persons (the principals) engage another person (the agent) to perform some service on their behalf, which involves delegating some decision-making authority to the agent. If the contractual parties to the relationship are utility maximizers there is good reason to believe that the agent will not always act in the best interests of the principals. The principals can limit divergences from their interests by establishing appropriate incentives for the agent and by incurring monitoring costs designed to limit the aberrant activities of the agent. In addition, in some situations it will pay the agent to expend resources (e.g., economic bonding costs) to guarantee that this agent will not take certain actions that would harm the principals or to ensure that the principals will be compensated if the agent does take such actions.

In most agency relationships the principals and agent will incur positive monitoring and economic bonding costs (non-pecuniary as well as pecuniary), and in addition there will be some divergence between the agent's decision and those decisions that would maximize the economic welfare of the principal. The dollar equivalent of the reduction in economic welfare experienced by the principals due to this divergence is also a cost of the agency relationship, and is referred to as the residual loss. *Jensen and Meckling define agency costs as: (1) the monitoring costs by the principal, the economic bonding expenditures by the agent; and (3) the residual loss.*

Jensen and Meckling (1976) note that since the relationship between stockholders and the manager of a corporation fit the definition of a pure agency relationship, the issues associated with the “separation of ownership and control” in the modern diffuse ownership corporation are intimately associated with the general problem of agency. Jensen and Meckling (1976) also point out that the problem of inducing an agent to behave as if this agent were maximizing the principals’ welfare is quite general. Agency exists in all organizations and in all cooperative efforts – at every level of management in firms, in universities, in mutual companies, in cooperatives, in governmental authorities and bureaus, in unions, and so forth. Jensen and Meckling (1976) focus on the analysis of agency costs generated by the contractual arrangement between the owners and top management of the corporation.

Jensen and Meckling (1976) argue that contractual relations are the essence of the firm, not only with employees (Alchian and Demsetz, 1972) but also with suppliers, customers, creditors and so on. The problem of agency costs and monitoring costs exists for all these contracts. Jensen and Meckling (1976) submit that most organizations serve as a nexus for a set of contracting relationships among individuals.

Jensen and Meckling (1976) argue that agency costs (i.e., monitoring costs, economic bonding costs, and the residual loss) are an unavoidable result of the agency relationship. *While Jensen and Meckling (1976) argue that agency costs are non-zero, these costs are not regarded as non-optimal in their framework.* In fact, Jensen and Meckling (1976) posit that since agency costs are borne entirely by the decision-maker, the decision-maker has the incentive to see that agency costs are minimized (because the decision-maker captures the benefits from the reduction in agency costs).

Jensen and Meckling (1976) conclude that agency costs are as real as any other costs. The level of agency cost depends among other things on statutory and common law, and human creativity in devising better contracts. Both the law and the sophistication of contracts relevant to the modern corporation are the products of an historical process in which there were strong economic incentives for individuals to minimize agency costs.

Having completed our survey of the agency literature, we now turn to the fifth chapter on resource-based theory, dynamic capabilities and real options. Resource-based theory is the dominant economic perspective in the research area of strategic management. I submit that students studying the economics of organization who want to be a first-rate scholars in this research area benefit greatly by appreciating the contributions of organization theory (e.g., the behavioral theory of the firm) and economics (e.g., transaction costs, property rights and agency theory). The next decade should see the development of resource-based theory by a new generation of enthusiastic and gifted scholars.

We begin chapter 5 with the seminal book by Edith Penrose (1959) on *The Theory of the Growth of the Firm*. This book had moderate success in the industrial organization research literature over the years (Scherer and Ross, 1990), and has had a major impact on the strategic management field from the mid-1980s up to the present.

## **Chapter 5    Resource-based Theory, Dynamic Capabilities and Real Options**

While early contributions to resource-based theory and dynamic capabilities came from the discipline of economics (e.g., Demsetz, 1973; Gort, 1962; Marris, 1964; Penrose, 1959; Richardson, 1960, 1972; Rubin, 1973; Slater, 1980), the past two decades (i.e., the 1983 – 2003 period) have seen significant contributions to resource-based theory and dynamic capabilities from the business field of strategic management (e.g., Foss, 1997; Heene and Sanchez, 1997; Volberda and Elfring, 2001). Logic dictates that (organizational) economic theory will continue to play an important role in the study of economic value creation and sustainable competitive advantage. After all, sustainable competitive advantage requires an understanding of market frictions, and there is a large and well-developed economics research literature on market failures that students studying the economics of organization can draw upon.

While the market-failures literature is well developed, the organizational-failures literature is comparatively less developed, thereby, providing research opportunities for students studying the economics of organization. Furthermore, resource-based theory/dynamic capabilities and real options research may develop into a paradigmatic approach to strategic management that contributes an important part to the evolving science of organization. Clearly, there is need for rigorous empirical research to establish both the nature of dynamic capabilities and the impact of dynamic capabilities on sustainable competitive advantage. Capabilities that can prove especially useful in dynamic business environments are operational and strategic flexibility.

We begin with the seminal work of Penrose (1959). Penrose (1959) provides (1) a general theory of the growth of the firm; (2) a theory of entrepreneurship based on the “subjective opportunity set” of the firm; (3) expansion based on indivisibility and the “balance of processes;” (4) a resource-based theory of diversification; and (5) a theory of expansion through acquisition and merger. In addition, Penrose (1959) provides a theory of the *limits* to the rate of the growth of the firm. In particular, Penrose (1959) argues that the binding constraint on the rate of the growth of the firm is provided by the capacities of its existing management --- the so-called “Penrose effect.”

Chandler’s (1990) *Scale and Scope* represents a culmination of a long quest by this business historian to chart the evolution of modern industrial enterprise. The book provides the reader with an extraordinary breadth and depth of knowledge concerning the development of managerial capitalism. The essence of successful firm strategy, Chandler (1990) argues, was the making of three interrelated investments: (1) investment in production to achieve the cost advantages of scale and scope; (2) investment in product-specific marketing, distribution, and purchasing networks; and (3) investment in managerial talent and management structure to plan, coordinate, and monitor the firm’s often dispersed operations. Chandler (1990) argues that such three-pronged investment enabled firms to develop organizational capabilities, which, in turn, provided the dynamic for the continuing growth of the enterprise.

Itami and Roehl (1987) emphasize the dynamic fit between resources and the environment. Itami and Roehl (1987) build on the work of Penrose (1959) concerning corporate growth and move the arguments forward by emphasizing the role of invisible assets of a firm, which are based on information. Invisible assets would include: intellectual property rights of

patents and trademarks, trade secrets, proprietary data files, personal and organizational networks, reputation and culture. Itami and Roehl (1987) argue that these invisible assets are often a firm's only real source of sustainable competitive advantage.

Nelson and Winter (1982) consider the promise and the problems of evolutionary modeling of economic change. Among the many benefits that may be derived from Nelson and Winter's (1982) theoretical approach that reconciles economic analysis with real-world business firm decision-making, the most important relate to improved understanding of technological change and the dynamics of the competitive process. Nelson and Winter's (1982) evolutionary theory is intrinsically dynamic theory, in which the heterogeneity of firms is a key feature.

This chapter on dynamic resource-based theory concludes with a research book that some in the strategic management field may find to be a curious choice. Over time, however, I anticipate that it will become abundantly clear that a key category in developing dynamic capabilities will involve strategies that enhance adaptability and strategic flexibility. Viewed in this light, Trigeorgis's (1996) research book is an important contribution to the dynamic capabilities research literature. Real options research has the potential to make a significant difference to our understanding of resource accumulation and capability-building processes and investment decision-making under uncertainty. Finally, supplementing the real options analysis with game-theoretic tools that capture competitive dynamics is promising for future research by students pursuing the evolving science of organization.

Penrose, Edith T. (1959). The Theory of the Growth of the Firm. New York: John Wiley & Sons.

Penrose (1959) is concerned with the growth of firms, and only incidentally with the size of the firm. Penrose (1959) argues that firm size is but a by-product of the process of growth, and that there is no "optimum," or even most profitable, size of the firm. Penrose (1959) is primarily concerned with a theoretical analysis of the growth process of the firm.

Penrose (1959) emphasizes the internal resources of a firm -- on the productive services available to a firm from its own resources, particularly the productive services available from management with experience within the firm. The (firm-specific) experience of management affects the productive services that all its other resources are capable of rendering. As management tries to make the best use of the resources available, a "dynamic" interacting process occurs that encourages continuous growth but limits the rate of growth of the firm. In order to focus attention on the crucial role of the firm's "inherited" resources, the environment is treated, in the first instance, as an "image" in the entrepreneur's mind of the possibilities and restrictions with which it is confronted. For it is, after all, such an "image" which in fact determines a person's behavior. Whether experience confirms expectations is another story.

The Firm in Theory. Penrose (1959) notes that in a private enterprise industrial economy the business firm is the basic unit for the organization of production. Because of its complexity and diversity, a firm can be approached with many different types of analysis -- sociological, organizational, engineering, or economic -- and from whatever point of view that seems appropriate to the business problem at hand.

The "theory of the firm" -- as it is called in the neoclassical economics literature -- was constructed for the purpose of assisting in the theoretical investigation of one of the central



problems of economic analysis -- the way in which prices and the allocation of resources among different uses are determined. It is but part of the wider theory of economic value. The "equilibrium" of the "firm" is, in essence, the "equilibrium output." As Boulding (1950:24) notes: "The firm is a strange bloodless creature without a balance sheet, without any visible capital structure, without debts, and engaged apparently in the simultaneous purchase of inputs and the sale of outputs at constant rates."

Penrose (1959) points out that if we become interested in other aspects of the firm then we ask questions that the "theory of the firm" is not designed to answer. Penrose (1959) wants to deal with the firm as a growing organization, not as a "price-and-output decision-maker" for given products. Penrose (1959) argues that the essential difference between economic activity inside the firm and economic activity in the "market" is that economic activity in the firm is carried on within an administrative organization (see Simon, 1947) while economic activity in the market is not. Penrose (1959) refers to this "court of last resort" in the firm as central management. It is the area of co-ordination -- the area of "authoritative communication" (Barnard, 1938; Simon 1947) — which defines the boundaries of the firm, and, consequently, it is a firm's ability to maintain sufficient administrative co-ordination to satisfy the definition of an industrial firm that sets the limit to its size as an industrial enterprise.

The Firm as a Collection of Productive Resources. According to Penrose (1959), a firm is more than an administrative unit; a firm is also a collection of productive resources where the choice of different uses of these resources over time is determined by administrative decision. The physical resources of a firm consist of tangible things -- plant, equipment, land and natural resources, raw materials, semi-finished goods, waste products and by-products, and even unsold

stocks of finished goods. There are also human resources available in a firm -- unskilled and skilled labor, clerical, administrative, financial, legal, technical, and managerial staff.

Penrose (1959) argues that strictly speaking, it is never *resources* themselves that are the "inputs" in the production process, but only the *services* that the resource can render. Resources consist of a bundle of potential services and can, for the most part, be defined independently of their use, while services cannot be so defined, the very word "service" implying a function, an activity. *It is largely in this distinction that we find the source of the uniqueness of each individual firm.*

The business firm as Penrose (1959) defines it, is both an administrative organization and a collection of productive resources. The general purpose of the business firm is to organize the use of its "own" resources together with other resources acquired from outside the firm for the production and sale of goods and services at a profit.

The term "entrepreneur" refers to individuals or groups within the firm providing the entrepreneurial services, whatever their position or occupational classification may be. Entrepreneurial services are those contributions to the operations of a firm that relate to the introduction and acceptance on behalf of the firm of new ideas, particularly with respect to products, location, and significant changes in technology, to the acquisition of new managerial personnel, to fundamental changes in the organization of the firm, to the raising of capital, and to the making of plans for expansion, including the strategic choice of method of expansion.

Penrose (1959) submits that a versatile type of executive service is needed if expansion requires major efforts on the part of the firm to develop new markets or entails branching out into new lines of production. Here, the imaginative effort, the sense of timing, the instinctive

recognition of what will catch on, or how to make it catch on, become of overwhelming importance. These services are not likely to be equally available to all firms. Firms not only alter the environmental conditions necessary for the success of their actions, but, even more importantly, they know that they can alter these conditions and that the environment is not independent of their own activities.

Expansion without Merger: The Receding Managerial Limit. Penrose (1959) notes three classes of explanation of why there may be a limit to the growth of firms --- managerial ability, product or factor markets, and uncertainty and risk -- the first explanation refers to conditions within the firm, the second explanation refers to conditions outside the firm, and the third explanation is a combination of internal attitudes and external conditions. The capacities of the *existing* managerial personnel of the firm necessarily set a limit to the expansion of that firm in any given period of time, and such management possessing firm-specific abilities cannot be hired in the marketplace.

Penrose (1959) argues that an administrative group is something more than a collection of individuals; an administrative group is a collection of individuals who have had experience in working together. For it is only in working together that "teamwork" can be developed. Experiences these individuals gain from working within the firm, and with each other, enables them to provide services that are uniquely valuable for the operations of the particular group with which they are associated. Existing management limits the amount of new management that can be hired at any point in time (after all, the services of existing management are required to instruct the new personnel).

Penrose (1959) submits that if a firm expands its organization more rapidly than the individuals in the expanding organization can obtain the experience with each other that is

necessary for the effective operation of the group, the efficiency of the firm will suffer. Since the services from current managerial resources control the amount of new managerial resources that can be absorbed, they create a fundamental limit to the amount of expansion a firm can undertake at any point in time. The amount of activity that can be planned at a given time period limits the amount of new personnel that can be profitably absorbed in the "next period." Over subsequent years, this idea has come to be known as the "Penrose effect." Moreover, as plans are completed and put into operation, managerial services absorbed in the planning processes will be gradually released and become available for further planning.

Penrose (1959) argues that knowledge comes to people in two different ways: (1) knowledge can be formally taught; and (2) knowledge can be achieved via learning-by-doing in the form of personal experience. Experience produces increased knowledge and contributes to "objective" knowledge in so far as its results can be transmitted to others. But experience itself can never be transmitted; experience produces a change -- frequently a subtle change -- in individuals and cannot be separated from them.

Increasing experience shows itself in two ways -- changes in knowledge acquired and changes in the ability to use knowledge. There is no sharp distinction between these two forms because to a considerable extent the ability to use old knowledge is dependent on the acquisition of new knowledge. But it is not exclusively so dependent; with experience a person may gain in wisdom, in sureness of movement, in confidence -- all of these become part of his or her very nature, and they are all qualities that are relevant to the kind and amount of services a person can give to the firm. Much of the experiences of business personnel are frequently so closely

associated with a particular set of external circumstances that a large part of a personnel's most valuable services may be available only under these circumstances.

A person whose past productive activity has been spent within a particular firm, for example, can, because of his or her intimate knowledge of the resources, structure, history, operations, and personnel of the firm render services to that firm which that person could give to no other firm without acquiring additional experience. Penrose (1959) submits that once it is recognized that the very processes of operating and of expanding are intimately associated with a process by which knowledge is increased, it becomes clear that the productive opportunity of a firm will change even in the absence of any change in external circumstances or in fundamental technological knowledge. New opportunities will open up that did not exist at the time expansion plans were made. That is, *the subjective opportunity set of the firm* will change.

Penrose (1959) hastens to add that not only is management the source for expansion but also management is a brake on expansion. A firm has a given amount of experienced managerial services available at any one time. Parts of these managerial services are needed for ordinary operation, and the rest of these managerial services are available for planning and executing expansion programs. The effect of uncertainty is to require that some of these available services be used to gather information, process the information, and reach conclusions about the possibilities of action in which the management team has confidence.

"Inherited" Resources and the Direction of Expansion. Penrose (1959) maintains that the external inducements to expansion include growing demand for particular products, changes in technology that call for production on a larger scale than before, discoveries and inventions the utilization of which seem particularly promising, and opportunities to obtain a better market position. Inducements to expansion also include backward integration to control sources of

supply, diversification of final products to spread risk, or expansion of existing or allied products to preclude the entry of new competitors. External obstacles to expansion include keen competition in markets for particular products that makes profitable entry or expansion in those markets difficult.

Penrose (1959) argues that while external inducements and obstacles have been widely discussed, little attention has been paid, in a systematic way at least, to the equally important internal influences on the direction of expansion. Internal obstacles arise when some of the important types of specialized services required for expansion in particular directions and are not available in sufficient amounts within the firm. In particular, internal obstacles arise when not enough of the managerial capacity and the technical skills required for the planning, execution, and efficient operation of a new program can be obtained from among existing experienced personnel. Internal inducements to expansion arise largely from the existence of a pool of unused productive services, resources, and specialized knowledge, all of which will always be found within any firm. To Penrose (1959), a resource can be viewed as a bundle of possible services. As long as resources are not used fully in current operations, there is an economic incentive for a firm to find a way of using them more fully.

Penrose (1959) maintains that three significant obstacles preclude the attainment of a "state of rest:"

1. Those arising from the familiar difficulties posed by the indivisibility of resources, which Penrose (1959: 69) calls "the jig-saw puzzle;"
2. Those arising from the fact that the same resources can be used differently under different circumstances; and
3. Those arising because in the ordinary processes of operation and expansion, new productive services are continually being created.

Penrose (1959) then argues how the division of labor (specialization) can lead to the growth of the firm and diversification. This process has been called the "virtuous circle" in which specialization leads to higher common multiples, and higher common multiples lead to greater specialization. Penrose (1959) also argues that diversification strategy can be driven by the desire to achieve multi-product economies of scale (which in modern strategic management language is called "economies of scope" (Teece, 1980)).

Penrose (1959) observes that for many purposes it is possible to deal with rather broad categories of resources, overlooking the lack of homogeneity in the members of the category. Economists usually recognize this simplification, stating that for convenience alone resources are grouped under a few headings -- for example, land, labor and capital -- but Penrose (1959) points out that the sub-division of resources may proceed as far as is useful, and according to whatever principles are most applicable for the business problem at hand. It is the heterogeneity of the productive services available or potentially available from its resources that give each firm its unique character.

Furthermore, the possibilities of using services change with changes in knowledge. Consequently, there is a close connection between the type of knowledge possessed by the personnel of the firm and the services obtainable from its material resources. The firm, then, is viewed as a collection of resources. Unused productive services shape the scope and direction of the search for knowledge. Knowledge and an economic incentive to search for new knowledge are "built into" the very nature of firms possessing entrepreneurial resources of even average initiative. Physically describable resources are purchased in the strategic factor markets for their known services; but as soon as these resources become part of a firm the range of services they

are capable of yielding starts to change. The services that resources will yield depend on the capacities of the people using them, but the development of the capacities of people is partly shaped by the resources they deal with. The two together create the distinctive, subjective, productive opportunity set of a particular firm.

If resources were completely non-specific, a firm could in principle produce anything. The selection of the relevant product-markets is necessarily determined by the "inherited" resources of the firm -- the productive services it already has. To be sure, the anticipation of consumer acceptance is a necessary condition of entrepreneurial interest in any product, but the original economic incentive to a great deal of innovation can be found in a firm's desire to use its existing resources more efficiently. There is a close relation between the various kinds of resources with which a firm works and the development of ideas, experience, and knowledge of its managers and entrepreneurs. Changing experience and knowledge affect not only the productive services available from resources, but also "demand" as seen by the firm.

Penrose (1959) further elaborates noting that unused productive services are, for the enterprising firm, at the same time a challenge to innovate, an economic incentive to expand, and a source of sustainable competitive advantage. Unused productive services facilitate the introduction of new combinations of resources -- innovations -- within the firm. Unused productive services are a selective force in determining the direction of expansion. Therefore, analysis is required of internal and external inducements, and internal and external obstacles for expansion.

The Economies of Diversification. Penrose (1959) argues that of all the outstanding characteristics of business firms perhaps the most inadequately treated in economic analysis is



the diversification of their activities. *Anticipating Teece (1982), Penrose (1959) argues that market imperfections are an important explanation of diversification strategy.* Diversification that involves a departure from the firm's existing areas may be one of three kinds:

- The entry into new markets with new products using the same production base;
- Expansion in the same market with new products based on a different area of technology; and
- Entry into new markets based on a different area of technology.

Penrose (1959) observes that a firm's opportunities are necessarily widened when the firm develops a specialized knowledge of a technology that is not very specific to any particular kind of product, for example, knowledge of different types of engineering or industrial chemistry. Diversification and expansion based primarily on a high degree of competence and technical knowledge in specialized areas of manufacture are characteristic of many of the largest firms in the economy. Penrose (1959) submits that this type of competence, together with the market position such competence and technical knowledge ensures, is the strongest and most enduring position a firm can develop.

Diversification through both internal and external expansion is likely to be extensive because of the wide variety of productive services generated within such firms, and because the competitive advantages these firms possess are conducive to expansion. Opportunities for expansion both within existing resource bases and through the establishment of new resource bases are likely to be so prevalent that the firm has carefully to choose among many different courses of action.

The Firm as a Pool of Resources. Penrose's (1959) thesis is that a firm is essentially a pool of resources the utilization of which is organized in an administrative framework. In a sense, the final products being produced by a firm at any given time represent one of several

ways in which the firm could be using its resources, an incident in the development of its basic potentialities. The continual change in the productive services and knowledge within a firm along with the continual change in external circumstances present the firm with a continually changing productive opportunity set.

The Rate of Growth of a Firm Through Time. Penrose (1959) notes that markets and firms are interacting institutions, each being necessary to the existence of the other. Penrose (1959) emphasizes that one of the more significant characteristics of entrepreneurial and managerial services is their heterogeneity, their uniqueness for every individual firm. The factors determining the availability of managerial services and the need for these services in expansion determines the maximum rate of growth of the firm, where rate of growth is defined as the percentage rate at which the size of the firm increases per unit of time. The possibility of acquiring other firms raises enormously the maximum rate of expansion, primarily because acquisition substantially reduces the managerial services required per unit of expansion.

Concluding Comments. Kor and Mahoney (2000) suggest the following list of key ideas that are derived from Penrose (1959):

- Firm growth can be usefully studied as a dynamic process of management interacting with resources.
- Firms are institutions created by people to serve the purposes of people.
- Services of resources are drivers of firm heterogeneity.
- Services that material resources will yield depend upon the knowledge possessed by human resources. The two together create a subjective opportunity set that is unique for each firm.
- Firm growth is a function of firm-specific experiences in teams.
- Managerial capability is the binding constraint that limits the growth rate of the firm – the so-called “Penrose effect.”
- Excess capacity of productive services of resources is a driver of firm growth.
- Unused productive services of resources can be a source of innovation.

- Firm diversification is often based on a firm's competencies that can lead to a sustainable competitive advantage.
- An important component of the competitive process is experimentation.

Finally, it is noted that some criticize Penrose's (1959) resources approach for ignoring the business environment. Penrose (1959), in fact, addresses this issue. Penrose (1959: 217) argues that whether or not we should treat the resources of the firm or its "environment" as the more important factor explaining growth depends on the question we ask: if we want to explain why different firms see the same environment differently, why some firms grow and some do not, or, to put it differently, why the environment is different for every firm, we must take the *resources approach*; if we want to explain why a particular firm or group of firms, with specified resources grows in the way it does we must examine the opportunities for the use of those resources. Penrose (1959) calls these opportunities for small firms the *interstices* in the economy. The productive opportunities of small firms are thus composed of those interstices left open by the large firms that the small firms see and believe they can take advantage of. Penrose (1959) concludes that management's experiments with different types of corporate structures *are in themselves an important aspect of competition*.

In my judgment, Penrose (1959) is the seminal work in resource-based theory that anticipates the works of Chandler (1962, 1990), which document organizational innovations and organizational capabilities that, in turn, provided an internal dynamic for the continuing growth of the modern industrial enterprise.<sup>10</sup>

---

<sup>10</sup> For further readings see Penrose (1955, 1960). Penrose (1960) provides a case study of the Hercules Powder Company to illustrate that growth is governed by a creative and dynamic interaction between a firm's productive resources and its market opportunities. Richardson (1972), Rubin (1973), Slater (1980), and Teece (1982) are influential journal articles in the economics research literature that build on Penrose (1959). For a recent assessment of Penrose (1959), see Kor and Mahoney (2000), which focuses on: (1) the research process that lead to Penrose's (1959) classic; (2) Penrose's (1959) contributions to the

We turn, then, to Chandler (1990). Chandler (1990) provides a detailed but highly generalized description and analysis of the creation and dynamic evolution of the central institution of managerial capitalism --- the modern industrial enterprise. These concepts and generalizations are then used to develop an explanatory theory of the evolution of the modern industrial enterprise. The richness of information provided in this research book can be helpful for students in the evolving science of organization in answering questions that have long concerned economists and business historians --- questions about changes in internal organization and management, competition and cooperation among firms; growth through horizontal acquisitions and mergers, vertical integration, expansion into international markets, diversification into new product lines; and finally, questions on how firm growth and economic performance are affected by legal requirements, government rulings, and cultural values.

---

discipline of strategic management; (3) the generative nature of Penrose's (1959) research for current resource-based theory; and (4) suggested future research building on Penrose's (1959) *resource approach*. In addition, Kor and Mahoney (2004) and Rugman and Verbeke (2002) consider Penrose's (1959) direct and indirect contributions to the modern resource-based view within strategic management. Pettus (2001) studies the Penrose effect in the deregulated trucking industry, while Tan (2003) and Tan and Mahoney (2004) provide empirical tests of Penrose (1959) in the context of multinational firms. Finally, Pitelis (2002) provides an excellent edited collection of recent writings that document the legacy of Penrose (1959) on contemporary research on the growth of the firm.

Chandler, Alfred D. (1990). Scale and Scope: The Dynamics of Capitalism. Cambridge, MA: Harvard University Press.

Chandler (1990) observes that in the last half of the nineteenth century a new form of capitalism appeared in the United States and Europe. The building and operating of rail and telegraph systems called for the creation of a new type of business enterprise. The massive investment required in constructing those systems, and the complexities of their operations, brought the separation of ownership and management. The enlarged enterprises came to be operated by teams of salaried managers who had little or no equity in the firm.

Chandler (1990) notes that the new forms of transportation and communication, in turn, permitted the rise of modern mass marketing and modern mass production. The unprecedented increase in the volume of production, and in the number of transactions, led the entrepreneurs who established the new mass-producing and mass-distributing enterprises — like the railroad personnel before them — to recruit teams of salaried managers.

Chandler (1990) examines the beginning and growth of managerial capitalism globally, focusing on the history of its basic institution, the modern *industrialized* enterprise, in the world's three leading industrial nations. They each had been rural, agrarian, and commercial; they each became industrial and urban. That transformation, in turn, brought the most rapid economic growth in business history. At the center of the transformation were the United States, Great Britain, and Germany that accounted for just over two-thirds of the world's industrial output in 1870. The industrial sector grew significantly in the United States and Germany; in Great Britain the development was slower, but sustained. Further, whereas Great Britain experienced only a moderate change of employment structure after the 1880s, the United States,

and Germany to a lesser degree, showed a dramatic transformation from an agrarian to a modern economy in which almost half of the employment centered in industry.

Chandler (1990) maintains that as a result of the regularity, increased volume, and greater speed of the flows of goods and materials made possible by the new transportation and communication systems, new and improved processes of production developed that for the first time in business history achieved substantial economies of scale and scope. Large manufacturing works applying the new technologies could produce at lower costs than could the smaller manufacturing works.

*Chandler (1990) observes that in order to benefit from the cost advantages of these new, high-volume technologies of production, entrepreneurs had to make three sets of interrelated investments:*

- First was an investment in production facilities large enough to utilize a technology's potential economies of scale and scope;
- Second was an investment in a national and international marketing and distribution network so that the volume of sales might keep pace with the new volume of production; and
- Third was investment in management: Entrepreneurs had to recruit and train managers not only to administer the enlarged facilities and increased personnel in both production and distribution, but also to monitor and coordinate those two basic functional activities and to plan and allocate resources for future production and distribution.

Chandler (1990) submits that it was this *three-pronged investment in production, distribution, and management that brought the modern industrial enterprise into being*. The first entrepreneurs to create such enterprises acquired substantive competitive advantages. Their industries quickly became oligopolistic; that is, dominated by a small number of first movers. These first-mover firms, along with a few challengers that subsequently entered the industry, no

longer competed primarily on the basis of price. Instead, these firms competed through functional and strategic effectiveness. These firms did so *functionally* by improving their product, their marketing, their purchasing, and their labor relations, and these firms did so *strategically* by moving into growing markets more rapidly, and by divesting out of declining markets more quickly and effectively than did their competitors.

Such rivalry for market share and profitability honed the enterprise's functional and strategic capabilities. These organizational capabilities, in turn, provided an internal dynamic for the continuing growth of the enterprise. In particular, these organizational capabilities stimulated its owners and managers to expand into more distant markets in their own country, and then to become multinational by moving abroad. These organizational capabilities also encouraged the firm to diversify by developing products competitive in markets other than the firm's original market, and so to become a multi-product enterprise.

Scale, Scope, and Organizational Capabilities. Chandler (1990) argues that the modern industrial enterprise can be defined as a collection of operating units, each with its own specific facilities and personnel, whose combined resources and activities are coordinated, monitored, and allocated by a hierarchy of middle and top managers. This hierarchy makes the activities and operations of the whole enterprise more than the sum of its operating units. The manufacturing enterprises became multi-functional, multi-regional, and multi-product because the addition of new units permitted these enterprises to maintain a long-term rate of return on investment by reducing overall costs of production and distribution, by providing products that satisfied existing demands and by transferring facilities and capabilities to more profitable markets when economic returns were reduced by competition, changing technology, or altered market demand.

Chandler (1990) submits that whatever the initial motivation for its investment in new operating units, the modern industrial enterprise has rarely continued to grow or maintain its competitive position over an extended period of time unless the addition of new units (and to a lesser extent the elimination of old units) has actually permitted the “visible hand” of its managerial hierarchy to reduce costs, to improve functional efficiency in marketing and purchasing as well as production, to improve existing products and processes and to develop new ones, and to allocate resources to meet the challenges and opportunities of ever-changing technologies and markets. Such a process of growth has provided organizations with the internal dynamic that has enabled these organizations to maintain their position of dominance as markets and technologies have changed. Chandler (1990) further argues that it was the development of new technologies and the opening of new markets, which resulted in economies of scale and scope, and in reduced transaction costs, which made the large multi-unit enterprise come when it did, where it did, and in the way it did.

Chandler (1990) maintains that coordination demanded the constant attention of a managerial team or hierarchy. The *potential* economies of scale and scope are a function of the physical characteristics of the production facilities. However, the *actual* economies of scale and scope, as measured by throughput, are a function of organizational capabilities. The full fruition of economies of scale and scope depend on knowledge, skill, experience, and teamwork — on the organizational capabilities essential to utilize the full potential of technological processes. Further, in many instances Chandler (1990) finds that the first company to build a plant of minimum efficient scale, and to recruit the essential management team to enable the enterprise to reach its full potential often remained the leader in its industry for decades.



Chandler (1990) indicates that *organizational capabilities* included, in addition to the skills of middle and top management, those of lower management and the work force. Organizational capabilities also included the facilities for production and distribution acquired to utilize fully the economies of scale and scope. Such organization capabilities provided the economic profits that, in large part, financed the continuing growth of the enterprise. Highly product-specific and process-specific, these organizational capabilities affected, indeed often determined, the direction and pace of first movers and challengers, and of the industries and even the national economies in which they operated (Collis, 1994).

Chandler (1990) emphasizes that only if these facilities and organizational capabilities were carefully coordinated and integrated could the enterprise achieve the economies of scale and scope that were needed to compete in national and international markets, and to continue to grow. Middle managers not only had to develop and apply functional-specific and product-specific managerial skills, but they also had to train and motivate lower-level managers, and to coordinate, integrate, and evaluate their work. Such *organizational capabilities*, of course, had to be created, and once established, these capabilities had to be maintained. Their maintenance was as great a challenge as their creation, for facilities depreciate, individual skills atrophy, and organizational capabilities can diminish. Moreover, changing technologies and markets constantly make existing facilities, individual skills, and organizational capabilities obsolete. One of the more critical tasks of the top management team has always been to maintain these organizational capabilities and to integrate these facilities and skills into a coherent, unified organization — so that the whole becomes more than the sum of its parts. Such organizational capabilities, in turn, have provided the source – the dynamic — for the continuing growth of the

enterprise. Organizational capabilities have made possible the earnings that supplied much of the funding for such growth.

As Chandler (1990, 1992) repeatedly emphasizes, in the collective individual industries that are so aptly documented, the first movers' initial, interrelated, three-pronged investments in manufacturing, marketing, and management created powerful *barriers to entry* (see also Porter, 1980). Challengers had to make comparable (sunk cost) investments at a greater risk, precisely because the first movers had already learned the ways of the new processes of production, were already dominating the markets for the new or greatly improved products, and were already reaping substantial economic returns from their initial investments. As the first movers' functional and organizational capabilities were honed, the difficulties of entry by newcomers became even more formidable. In the sale of consumer products, particularly branded, packaged goods, barriers to entry were reinforced by advertising, vertical tying contracts, and exclusive franchising. In the more technologically advanced producer-goods industries, patents reinforced these entry barriers. In Europe, first movers strengthened their strategic positioning still further by arranging inter-firm agreements as to price, output, and market territories.

A New Era of Managerial Capitalism? Chandler (1990) notes that the historian who has studied the past experience of the business enterprise is in a better position than most analysts to identify current business practices that are truly new. Chandler (1990) observes that of the many recent changes in the growth, management, and financing of the modern industrial enterprise, six have no historical precedents. These changes include:

- The adaptation of a new corporate strategy of growth — that of moving into new markets where the organizational capabilities of the enterprise do not provide competitive advantages;

- The separation of top management in the corporate office from middle management in the operating divisions;
- The extensive and continuing divestiture of operating units;
- The buying and selling of corporations as a distinct business in its own right;
- The role played by portfolio managers in the capital markets; and
- The evolution of those capital markets to facilitate the coming of what has been termed “a market for corporate control.”

Chandler (1990) concludes that his research book has only begun to map the evolution of the industrial enterprise in the United States, Great Britain, and Germany from the 1880s to the 1940s. Valid description and analysis on which generalizations can be made must await an in-depth, industry-by-industry, country-by-country historical study.<sup>11</sup> Much more work needs to be done that certainly may modify the patterns of institutional change that Chandler (1990) has outlined. Clearly, there are research opportunities for those students studying the economics of organization who combine the craft of the business historian and the analytical skills derived from the resource-based/dynamic capabilities perspective. Indeed, Chandler (1992) provides insights that connect (company-specific) organizational capabilities and the economics of organization.

To develop further the dynamic capabilities perspective, we turn next to Itami and Roehl’s (1987) contribution to dynamic capabilities theory. Itami and Roehl (1987) emphasize the role of (environmental, corporate and internal) information flow. Environmental information flow includes discovering customer preferences and maintaining competitor intelligence. Corporate information flow includes reputation, brand image, and marketing know-how. Internal information flow includes corporate culture and managerial capabilities (e.g., routines).

---

<sup>11</sup> Resource-based/dynamic capabilities theory has recently been empirically corroborated in the context of international business studies (e.g., Anand and DeLois, 2002; Anand and Singh, 1997). Peng (2001) documents the extent to which resource-based theory has diffused in international business research.

Itami, Hiroyuki and Thomas E. Roehl (1987). Mobilizing Invisible Assets. Cambridge, MA: Harvard University Press.

Itami and Roehl (1987) provide a strategic logic that is heavily influenced by Penrose (1959) and emphasizes the vital contribution of accumulated experience and information to a corporation's strategic resources. Itami and Roehl (1987) emphasize that the intangible assets, such as a particular technology, accumulated consumer information, brand name, reputation, and corporate culture, are invaluable to the firm's competitive advantage. In fact, these invisible resources are often a firm's only real source of competitive edge that can be sustained over time.

Itami and Roehl (1987) emphasize that current strategy, because it can change the level of invisible assets, is more than the basis for short-term competitive advantage, current strategy provides the foundation for future strategy, and adds to or erodes the invisible resource base. The competitive success of a strategy is dependent on the firm's invisible assets, but the dynamics of invisible assets (their accumulation and depreciation over time) is also largely determined by the content of that strategy. Itami and Roehl (1987) explore how invisible assets affect, and are affected by, the firm's strategy. Decisions made today can affect a firm's long-term capabilities and adaptability because such decisions often determine the accumulation of invisible assets.

Itami and Roehl (1987) maintain that many invisible resources are quite fixed. There is no easy way to obtain a well-known brand name or advanced technical production skills in the market. Nor can money buy an instantaneous change in corporate culture and employee morale. Accumulation of these invisible resources requires on-going, conscious, and time-consuming efforts; you cannot just go out and buy them off the shelf. For this reason, a firm can

differentiate itself from competitors through its invisible resources. If a resource can be bought, competitors with sufficient financial resources can gain access to it. And if a resource can be created quickly, competitors will have ready access to such a resource through imitation. But competitors cannot do this easily with invisible resources.

The important features of invisible resources are: (1) these resources are unattainable with money alone; (2) they are time consuming to develop; (3) they are capable of multiple simultaneous use; (4) these resources yield multiple, simultaneous benefits. These features of invisible resources make it crucial to consider carefully strategies for accumulating invisible resources.

*Information is at the heart of invisible resources.* Information-based invisible resources include not only the stock of accumulated information in the firm, but also include the channels that handle the flow of information of importance to the firm. Information can be classified as environmental, corporate, or internal.

*Environmental information* flows from the environment to the firm, creating invisible assets related to the environment. This type of information flow includes production skills, customer information, and channels for bringing in information.

*Corporate information* flows from the firm to the environment, creating invisible assets stored in the environment. This category of information flow includes such invisible assets as corporate reputation, brand image, corporate image, and influence over the distribution and its parts suppliers, as well as marketing know-how.

*Internal information* originates and terminates within the firm, again affecting the invisible asset stock. This category of information flow includes corporate culture, morale of workers, and management capabilities, as well as the firm's ability to manage information, the

employees' ability to transmit and use the information in decision making, and the employees' habits and norms of effort expended. *Successful accumulation of invisible resources comes down to control of the information flow.*

In my judgment, Itami and Roehl (1987) is a seminal contribution to resource-based theory and the dynamic capability approach. Invisible assets serve as a focal point of strategy development and growth. Students studying the economics of organization are served well in examining closely this often over-looked classic.

We turn next to a classic that almost everyone recognizes as the seminal and path-breaking book on evolutionary economics and dynamic capabilities. Nelson and Winter (1982) provide a wealth of strategic issues for consideration by current students who want to contribute to the evolving science of organization. Nelson and Winter (1982) provide the organization-theoretic foundations of economic evolutionary theory --- the building blocks of this evolutionary theory include individual skills and organizational capabilities. Nelson and Winter (1982) develop an evolutionary model of economic growth and a perspective that emphasizes the role of Schumpeterian competition.

Nelson, Richard R., and Sidney G. Winter (1982). An Evolutionary Theory of Economic Change. Cambridge, Mass.: Harvard University Press.

*Nelson and Winter (1982) argue that much of firm behavior can be more readily understood as a reflection of general routines and strategic orientations coming from the firm's past than as the result of a detailed survey of the remote twigs of a decision tree extending into the future.* Nelson and Winter (1982) acknowledge their intellectual debts to Joseph Schumpeter and Herbert Simon. Schumpeter (1934, 1950) points out the right problem -- how to understand economic change -- and Schumpeter's vision encompassed many of the important elements of the answer. Simon (1982) provides a number of specific insights into human and organizational behavior that are reflected in Nelson and Winter's (1982) theoretical models. But, most importantly, Simon's (1947, 1982) works encourage Nelson and Winter (1982) in maintaining the view that there is much more to be said on the problem of rational behavior in the world of experience than cannot be adequately stated in the language of orthodox economic theory.

Nelson and Winter (1982) develop an evolutionary theory of the organizational capabilities and behaviors of business firms operating in a market environment. The firms in their evolutionary theory are treated as motivated by profitability and engaged in search for ways to improve their profitability, but the firm's actions are not assumed to be profit maximizing over well-defined and exogenously given choice sets. Evolutionary theory emphasizes the tendency for the more profitable firms to drive the less profitable firms out of business. However, Nelson and Winter (1982) do not focus their analysis on hypothetical states of "industry equilibrium," in which all the unprofitable firms are no longer in the industry and the profitable firms are at their desired size.

Relatedly, the modeling approach employed in Nelson and Winter (1982) does not use the familiar maximization calculus to derive equations characterizing the behavior of firms. Rather, firms are modeled as having, at any given time, certain organizational capabilities and decision rules. Over time, these organizational capabilities and decision rules are modified as a result of both deliberate problem-solving efforts and random events. And over time, the economic analogue of natural selection operates as the market determines which firms are profitable and which are unprofitable, and tends to winnow out the unprofitable firms. Supporting Nelson and Winter's (1982) analytical emphasis on this sort of evolution by natural selection is a view of "organizational genetics" --- the processes by which traits of organizations, including those traits underlying the capability to produce output and to make profits, are transmitted through time.

Nelson and Winter (1982) give attention to uncertainty, bounded rationality, the presence of large corporations, institutional complexity, and the dynamics of the actual adjustment process. Considerable attention is also given to imperfect information and imperfect competition, to transaction costs, indivisibilities, increasing returns, and historical change.

Although Nelson and Winter (1982) stress the importance of certain elements of continuity in the economic process, Nelson and Winter (1982) do not deny (nor does contemporary biology deny) that change is sometimes rapid. Also, some people who are particularly alert to teleological fallacies in the interpretation of biological evolution seem to insist on a sharp distinction between explanations that feature the processes of "blind" evolution and those that feature "deliberate" goal-seeking. Whatever the merits of this distinction in the context of the theory of biological evolution, such a distinction is unhelpful and distracting in the



context of Nelson and Winter's (1982) theory of the business firm. It is neither difficult nor implausible to develop models of firm behavior that interweave "blind" and "deliberate" processes. Indeed, in human problem solving itself, both elements are involved and difficult to disentangle. Relatedly, Nelson and Winter (1982) describe their theory as unabashedly Lamarckian: the evolutionary economics theory of the firm contemplates both the "inheritance" of acquired characteristics and the timely appearance of variations under the stimulus of adversity.

Nelson and Winter's (1982) general term for all regular and predictable behavioral patterns of firms is "routine." Nelson and Winter (1982) use this general term to include characteristics of firms that range from well-specified technical routines for producing things, through procedures for hiring and firing, ordering new inventory, or stepping up production of items in high demand, to policies regarding investment, research and development (R&D), or advertising, and business strategies about product diversification and overseas investment. In Nelson and Winter's (1982) evolutionary theory, these routines play the role that genes play in biological evolutionary theory. They are a persistent feature of the organism and determine its possible behavior (though *actual* behavior is determined also by the environment).

Most of what is *regular and predictable* about business behavior is plausibly subsumed under the heading "routine." The fact that not all business behavior follows regular and predictable patterns is accommodated in evolutionary theory by recognizing that there are stochastic elements both in the determination of decisions and of decision outcomes. From the perspective of a participant in business decision-making, these stochastic elements may reflect the result of tumultuous meetings or of confrontations with complex problems under crisis conditions; but from the viewpoint of an external observer seeking to understand the dynamics of

the larger system, these phenomena are difficult to predict.

Whereas in orthodox theory, decision rules are assumed to be the consequence of maximization, in evolutionary theory decision rules are treated as reflecting at any moment in time the historically given routines governing the actions of a business firm. Routine-changing processes are modeled as "searches." Nelson and Winter's (1982) concept of search is the counterpart of that of mutation in biological evolutionary theory. Through the joint action of search and selection, the firms evolve over time, with the conditions of the industry in each period bearing the seeds of its condition in the following period.

Just as some orthodox microeconomic ideas seem to find their most natural mathematical expression in the calculus, the foregoing verbal account of economic evolution seems to translate naturally into a description of a Markov process -- though one in a rather complicated state space. The process is not deterministic; search outcomes, in particular, are partly stochastic. Thus, what the industry condition of a particular period really *determines* is the probability distribution of its condition in the following period.

Important antecedents of Nelson and Winter (1982) have been described in previous chapters:

- Behavioral Theory of the Firm (Cyert and March 1963; Simon, 1947);
- Transaction Costs Theory (Williamson, 1975);
- Theory of the Growth of the Firm (Penrose, 1959); and
- Business History (Chandler, 1962).

Chandler (1962) demonstrates that the organizational capabilities of a firm are embedded in its organizational structure, which is better adapted to certain strategies than to others. Thus,

strategies at any point in time are constrained by the organization. Also, a significant change in a firm's strategy is likely to call for a significant change in its organizational structure.

Nelson and Winter (1982) build on the concept of Schumpeterian competition. Schumpeter's credentials as a theorist of bounded rationality could hardly be more incisively established than in the following passage (1934: 80):

The assumption that conduct is prompt and rational is in all cases a fiction. But it proves to be sufficiently near to reality, if things have time to hammer logic into men. Where this has happened, one may rest content with this fiction and build theories upon it ... Outside of these limits our fiction loses its closeness to reality. To cling to it there also, as the traditional theory does, is to hide an essential thing and to ignore a fact which, in contrast with other deviations of our assumptions from reality, is theoretically important and the source of the explanation of phenomena which would not exist without it.

Nelson and Winter (1982) observe that a consistent theme in retrospective studies is that failure occurs not because the intelligence system failed to acquire warning signals but because the intelligence system failed to process, relate, and interpret those signals into a message relevant to available choices. Intelligence analysts and decision-makers have only a limited amount of time each day, limited communication channels to connect their systems, and limited assistance in the task of organizing, analyzing, and thinking about the available information. Sometimes, highly "obvious" and emphatic signals get lost in the noise as a result of these limitations. The events of 9/11/01 are a compelling recent example. Nelson and Winter (1982) see no reason to think that economic decision-making is any different in this regard.

There is similarly a fundamental difference between a situation in which a decision - maker is uncertain about the state of X, and a situation in which the decision-maker has not given any thought to whether X matters or not. To treat them calls for a theory of attention, not a theory that assumes that everything always is attended to, but that some things are given little

weight (for objective reasons). In short, the most complex models of maximizing choice do not come to grips with the problem of bounded rationality.

Skills. Nelson and Winter (1982) develop the basic postulates about behavior in evolutionary theory. Although evolutionary economics theory is concerned with the behavior of business firms and other organizations, Nelson and Winter (1982) find it useful to begin the analysis with a discussion of some aspects of individual behavior. An obvious reason for doing so is that the behavior of an organization is, in a limited but important sense, reducible to the behavior of the individuals who are members of that organization. Regularities of individual behavior must therefore be expected to have consequences, if not counterparts, at the organizational level (see Dosi, Nelson and Winter, 2000).

Nelson and Winter (1982) propose that individual skills are the analogue of organizational routines, and that an understanding of the role that routinization plays in organizational functioning is therefore obtainable by considering the role of skills in individual functioning. By a "skill" Nelson and Winter (1982) mean an ability to achieve a smooth sequence of coordinated behavior that is ordinarily effective relative to its objectives, given the context in which the skill normally occurs. Thus, the ability to serve a tennis ball is a skill, as is the ability to engage in competent carpentry, drive a car, operate a computer, set up and solve a linear programming model, or judge which job candidate to hire. Important characteristics of skills are:

- Skills are programmatic (i.e., a sequence of closely followed steps).
- The knowledge that enables a skillful performance is, in large measure, tacit knowledge, in the sense that the performer is not fully aware of the details of the performance, and finds it difficult or impossible to articulate a full account of those details; and

- The exercise of a skill often involves the making of numerous "choices" -- but to a considerable extent the options are selected automatically and without awareness that a choice is being made.

Nelson and Winter (1982) note that these three aspects of skilled behavior are closely interrelated. Skilled human performance is automatic in the sense that most of the details are executed without conscious volition. Indeed, a welcome precursor of success in an effort to acquire a new skill is the diminishing need to attend to details. Although "impressiveness" is obviously a matter of degree and relative to expectation, only the most unmoving can escape being impressed, at some point, by a skillful performance.

The late scientist-philosopher Michael Polanyi wrote extensively of the central place in the general scheme of human knowledge occupied by knowledge that cannot be articulated -- tacit knowledge. On the simple observation "We know more than we can tell," Polanyi (1962) built an entire philosophical system. Polanyi (1962) notes that to be able to do something, and at the same time be unable to explain how it is done, is more than a logical possibility -- it is a common situation. Polanyi, (1962:49) offers a good example early in the discussion of skills:

I shall take as my clue for this investigation the well-known fact that the aim of a skillful performance is achieved by the observance of a set of rules, which are not known as such to the person following them. For example, the decisive factor by which the swimmer keeps himself afloat is the manner by which he regulates his respiration; he keeps his buoyancy at an increased level by refraining from emptying his lungs when breathing out and by inflating them more than usual when breathing in; yet this is not generally known to swimmers.

Nelson and Winter (1982) note that the tacitness of a skill, or rather of the knowledge enabling a skill, is a matter of degree. Words are probably a more effective vehicle for communicating the skills of elementary algebra than for those of carpentry, and more effective for carpentry than for gymnastic stunts. Also, a trait that distinguishes a good instructor is the

ability to discover introspectively, and then articulate for the student, much of the knowledge that ordinarily remains tacit. Skill involves the observance of a set of rules, which are not known as such to the person following them.

What are some determinants of the degree of tacitness? There is, first of all, a limit imposed by the feasible time rate of information transfer through symbolic communication, which may be well below the rate necessary or appropriate in the actual performance.

A second consideration that limits the articulation of the knowledge underlying a skill is the limited causal depth of the knowledge. Polanyi's (1962) swimming example illustrates the point that the possession of a skill does not require theoretical understanding of the basis of the skill. Yet, this does not imply that an attempt to articulate the basis of the skill would not benefit from the availability of this terminology. Perhaps some novice swimmers would be helped by Polanyi's (1962) brief explanation of the body's buoyancy.

The third aspect of the limitation of articulation is the coherence aspect -- that of the whole versus the parts. Efforts to articulate "complete" knowledge of something by exhaustive attention to details and thorough discussion of preconditions succeed only in producing an incoherent message. This difficulty is probably rooted to a substantial extent in the related facts of the linear character of language-based communication, the serial character of the "central processor" of the human brain, and the relatively limited capacity of human short-term memory. Given these facts, the possibilities of articulating both the details and the coherent patterns they form -- the relationships among the details -- are necessarily limited. In short, much operational knowledge remains tacit because the knowledge cannot be articulated fast enough, because the knowledge is impossible to articulate all that is necessary to a successful performance, and

because language cannot simultaneously serve to describe relationships and characterize the things related.

The knowledge contained in the how-to-do-it book and its various supplements and analogues tends to be more adequate when the pace of the required performance is slow and pace variations are tolerable, where a standardized, controlled context for the performance is somehow assured, and where the performance as a whole is truly reducible to a set of simple parts that relate to one another in some very simple ways. To the extent that these conditions do not hold, the role of tacit knowledge in the performance may be expected to be large.

Finally, it should be emphasized that economic costs matter. Whether a particular bit of knowledge is in principle articulable or necessarily tacit is not the relevant question in most behavioral situations. Rather, the question is whether the economic costs associated with the obstacles to articulation are sufficiently high so that the knowledge, in fact, remains tacit.

There is in a sense a tradeoff between ability and deliberate choice, a tradeoff imposed ultimately by the fact that rationality is bounded. The advantages of skills are attained by suppressing deliberate choice, confining behavior to well-defined channels, and reducing option selection to just another part of the "program" (March and Simon, 1958).

Orthodox microeconomic theory treats the skillful behavior of the businessman as maximizing choice, and "choice" carries connotations of "deliberation." Nelson and Winter (1982), on the other hand, emphasize the automaticity of skillful behavior and the suppression of choice that this skillful behavior involves.

Organizational Capabilities and Behavior. The organizations that Nelson and Winter (1982) envisage are organizations that face a substantial coordination problem, typically because these organizations have many members, performing many distinct roles, who make

complementary contributions to the production of a relatively small range of goods and services.

Nelson and Winter (1982) provide several salient functions of routines:

1. *Routine as Organizational Memory.* The routinization of activity in an organization constitutes the most important form of storage of the organization's specific operational knowledge. Basically, Nelson and Winter (1982) claim that organizations remember by doing. Exercise of a routine serves as parsimonious organizational memory. Recall that Arrow (1974) has given particular emphasis to the internal dialectic or "code" of an organization as a key resource of the economies that formal organization provides, and as an important cause of persistent differences among organizations;
2. *Routine as Truce.* Routine operation involves a comprehensive truce in intra-organizational conflict (Cyert and March, 1963). Adaptations that appear "obvious" and "easy" to an external observer may be foreclosed because such adaptations involve a perceived threat to internal political stability; and
3. *Routine as Target: Control, Replication and Imitation.* Nelson and Winter (1982) note that replication is often a non-trivial exercise. Polanyi (1962:52) observes that:

The attempt to analyze scientifically the established arts has everywhere led to similar results. Indeed, even in modern industries the indefinable knowledge is still an essential part of technology. I have myself watched in Hungary a new, imported machine for blowing electric lamp bulbs, the exact counterpart of which was operating successfully in Germany, failing for a whole year to produce a single flawless bulb.

The assumption that perfect replication is possible in evolutionary models is intended primarily to reflect the advantages that favor the going concerns that attempt to do more of the same, as contrasted with the difficulties that they would encounter in doing something else or that others would encounter in trying to copy their success. There are some potential obstacles to replication that may be difficult to overcome even at very high cost. Some employees at the old plant may be exercising complex skills with large tacit components, acquired through years of experience in the firm. Others may have skills of lesser complexity and tacitness, but be poor at teaching those skills to someone else -- doing and teaching are, after all, different. Some members may, for various reasons, be unwilling to cooperate in the process of transferring their segment of the memory contents to someone else; they may, for example, be unwilling to disclose how easy their job really is, or the extent of the shortcuts they take in doing it. Williamson (1975) addresses the question of the incentives of organization members to disclose idiosyncratic information of importance to the organization's



functioning under the rubric "information impactedness." Nelson and Winter (1982) note that the target routine may involve so much idiosyncratic and "impacted" tacit knowledge that even successful replication is highly problematic, let alone imitation from a distance.

4. *Routines and Skills: Parallels.* Nelson and Winter (1982) note that routines are the skills of an organization. Organizations are poor at improvising coordinated responses to novel situations; an individual lacking skills appropriate to the situation may respond awkwardly, but an organization lacking appropriate routines may not respond at all.
5. *Optimal Routines and Optimization Routines.* The heart of Nelson and Winter's (1982) proposal is that the behavior of firms can be explained by the routines that these firms employ. Modeling the behavior of the firm means modeling the routines and how these firms change over time.
6. *Routines, Heuristics, and Innovation.* According to Nelson and Winter (1982), innovation involves change in routine. Similarly, Schumpeter (1934: 65-66) identified innovation with the "carrying out of new combinations." A "heuristic" is any principle or device that contributes to reduction in the average search to solution. Schumpeter (1950) proposed that sometime during the twentieth century the modern corporation "routinized innovation."
7. *Routines as Genes.* Nelson and Winter (1982) argue that as a first approximation, firms may be expected to behave in the future according to the routines they have employed in the past. Efforts to understand the functioning of industries and larger systems should come to grips with the fact that highly flexible adaptation to change is not likely to characterize the behavior of individual firms. Evolutionary theory does come to grips with this fact.

Static Selection Equilibrium. Nelson and Winter (1982) note that in Friedman (1953) there is no hint that an evolutionary theory is an alternative to orthodoxy. Rather, the proposition is that selection forces may be the proper explanation of why orthodox theory is a good predictive engine. Alchian (1950) sets forth a perspective regarding firm behavior that resembles Nelson and Winter's (1982) in many ways, stressing the element of luck in determining outcomes, the role of learning by trial and feedback and imitation in guiding firms to do better, and of selection forces in molding what firms and industries do. Alchian states that: "What really counts is the various actions actually tried, for it is from these that success is selected, not from

some set of perfect actions. The economist may be pushing his luck too far in arguing that actions in response to changes in environment and changes in satisfaction with the existing state of affairs will converge as a result of adaptation or adoption towards the optimum action that would have been selected if foresight had been perfect" (1950: 218). This statement is not an argument that selection forces provide a reason for adherence to orthodox theory, but rather a suggestion that there may be some important differences between an orthodox and an evolutionary perspective. Selection works on what exists, not on the full set of what is theoretically possible (Langlois, 1986; O'Driscoll and Rizzo, 1985).

Competition is viewed as a dynamic process involving uncertainty, struggle, and disequilibrium, not as a tranquil equilibrium state. In evolutionary theory, decision rules are viewed as a legacy from the firm's past and hence appropriate, at best, to the range of circumstances in which the firm customarily finds itself, and are viewed as unresponsive, or inappropriate, to novel situations or situations encountered irregularly.

The heart of the R&D-innovation problem is that reasonable people will disagree about what techniques will be best when. Importantly, this uncertainty is a major reason why it makes sense to have R&D largely conducted by competitive business firms who make their own entrepreneurial decisions, rather than place R&D decisions under more centralized control (see Nelson, 1996).

Dynamic Competition and Technical Progress. The market system is (in part) a device for conducting and evaluating experiments in economic behavior and organization. The meaning and merit of competition must be appraised accordingly. In Schumpeter's (1934) terms

competition involves "carrying out new combinations." Schumpeter's (1934) concept of innovation was a broad one. Schumpeter's five identified cases were (1934: 66):

- The introduction of a new good;
- The introduction of a new method of production;
- The opening of a new market;
- The opening of a new source of supply; and
- The carrying out of the new organization of any industry, like the creation of a monopoly position.

Although Schumpeter (1934) is particularly noteworthy for this emphasis on experimentation, most of the great economists, from Adam Smith (1776) to the onset of the modern period of formalization, gave some weight to the experimental role of competitive markets. An essential aspect of Schumpeterian competition is that firms do not know *ex ante* whether it pays to try to be an innovator or an imitator, or what levels of R&D expenditures might be appropriate. Only the course of events over time will determine and reveal what strategies are the better ones. And even the verdict of hindsight may be less than clear.

Normative Organizational Economics from an Evolutionary Perspective. Nelson and Winter (1982) note that the modern advocacy of private enterprise solutions tends to suffer from vagueness or utopianism in its treatment of institutional matters. Three particularly important (and closely interrelated) ones involve the treatment of property rights, contracts, and law enforcement. In almost all formalized economic theory, property rights and contractual obligations are assumed to be costlessly delineated in unambiguous terms, and enforcement of the civil and criminal law is perfect and costless. By virtue of the combined force of these assumptions of clarity, perfection, and zero transaction costs, the problem of providing the basic institutional underpinnings of a system of voluntary exchange is assumed away. It is then not

too surprising that voluntary exchange can be shown to be a largely effective economic solution to such problems as are left.

A legal system that could approach the theoretical standards of clarity and perfection in the delineation and enforcement of entitlements would be an elaborate and expensive system indeed. This is particularly obvious if the system of entitlements is supposed to be so sophisticated as to bring within its scope all of the externality problems that economists sometimes treat as "merely" problems in the definition and enforcement of property rights -- for example, the question of whether a chemical plant is entitled to dispose of its hazardous wastes in ways that contaminate the ground water, or whether neighboring property owners are entitled to uncontaminated ground water. If the anatomy of market failure is a function of institutional structure, institutional structure itself evolves in part in response to perceived problems with the status quo.

Nelson and Winter (1982) conclude that the attempt to optimize and accordingly to control technological advance will, according to evolutionary theory, lead not to efficiency but to inefficiency. In terms of empirical testing of evolutionary organizational economics, Nelson and Winter (1982) note that organizations that operate many very similar establishments -- for example, retailing and fast-food chains -- provide a natural laboratory for studying the problems of control and replication. Students with interest in the area of resources and organizational capabilities should see Foss (1997), Langlois and Robertson (1995), and Nelson and Winter (2002) for an update on recent research literature on dynamic capabilities and evolutionary economics.

## Theory and Applications:

### A Brief Review of Contemporary Resource-Based and Dynamic Capabilities Literature.

Resource-based theory addresses some of the fundamental issues in strategy (Rumelt, Schendel and Teece, 1994; Teece, 2000). Taking 1982 (when Nelson and Winter (1982) was published) as the starting point, we consider some seminal contributions to resource-based theory:

- Lippman and Rumelt (1982): Causal ambiguity inherent in the creation of productive processes is modeled by attaching an irreducible *ex ante* uncertainty to the level of firm efficiency that is achieved by sequential entrants. Without recourse to scale economies or market power, the model generates equilibria in which there are stable inter-firm differences in profitability. Sustainable competitive advantage results from the rich connections between uniqueness and causal ambiguity (see also, Reed and DeFillipi, 1990, and Rumelt, 1984).
- Teece (1982): This paper outlines a theory of the multi-product firm. Important building blocks include excess capacity and its creation, market imperfections, and the characteristics of organizational capabilities, including its fungible and tacit character. Teece (1982) both heavily acknowledges, and builds on, Penrose (1959). Teece (1982) argues that a firm's capabilities are upstream from the end product --- organizational capabilities might well find a variety of end product applications (as Penrose's (1960) case study of the Hercules Powder Company effectively shows).
- Wernerfelt (1984, 1995): Building on the seminal work of Penrose (1959), this work argues that strategy involves a balance between the utilization of existing resources and the development of new resources.
- Montgomery and Wernerfelt (1988): According to resource-based theory (Teece, 1982), firms diversify in response to excess capacity of resources that are subject to market frictions. By probing into the heterogeneity of these resources, this paper develops the corollary that firms that diversify most widely should expect the lowest average (Ricardian) rents. An empirical test, with Tobin's  $q$  as a measure of rents, is consistent with this resource-based theory.
- Dierickx and Cool (1989): This paper draws the distinction between tradeable and non-tradeable resources (e.g., reputation). The paper further argues for a time-based view of competitive strategy (due, in part, to time compression diseconomies).
- Cohen and Levinthal (1990): This paper argues that prior related knowledge confers an ability to recognize the economic value of new information, assimilate the information,

and apply the information to commercial uses. These dynamic capabilities constitute a firm's "absorptive capacity." Cross-sectional data on technological opportunity and appropriability conditions in the American manufacturing sector collected for R&D lab managers and the FTC Line-of-Business data indicate that R&D both generates innovation and facilitates learning.

- Henderson and Clark (1990): This paper distinguishes between the components of a product, and the ways that the components are integrated into the system that is the product "architecture." Data were collected during a two-year, field-based study of the photolithographic alignment equipment industry. The core of the data is a panel data set consisting of research and development costs and sales revenue by product for every product development project conducted between 1962, when the work on the first commercial product began and 1986. The concept of architectural innovation provides rich resource-based connections between innovation and organizational capabilities.
- Barney (1991): This often-cited paper suggests that the search for sources of sustainable competitive advantage must focus on resource heterogeneity and immobility. This paper argues that sustainable competitive advantage is derived from resources that are valuable, rare, imperfectly imitable (due to path-dependence, causal ambiguity, and social complexity), and non-substitutable.
- Chatterjee and Wernerfelt (1991): This paper theoretically and empirically investigates the resource-based view that firm's diversify, in part, to utilize excess productive resources. In particular, empirical evidence corroborates that excess physical resources, and most knowledge-based resources lead to more related diversification.
- Conner (1991): This paper analyzes resource-based theory as a new theory of the firm. The paper makes insightful connections between resource-based theory and Schumpeterian (1934, 1950) competition.
- Montgomery and Hariharan (1991): Using a sample of 366 firms in the FTC's Line-of-Business database, this research indicates that growth and diversification in large established firms result from a process of matching a firm's lumpy (indivisible) and ever-changing resources with dynamic market opportunities. Overall, this research provides empirical support for Penrose's (1959) theory of diversified entry: Unused productive services of resources are a selective force in determining the direction of firm-level expansion.
- Mosakowski (1991): Using a longitudinal data set, a sample of 86 entrepreneurial firms in the computer software industry that completed an IPO in 1984 is examined. Empirical

findings suggest that strategies that represent rare, inimitable and non-substitutable resources are a source of competitive advantage.

- Porter (1991): This paper argues that firms have accumulated differing resources because of differing strategies and configurations of (value-chain) activities. Resources and activities are, in a sense, duals of each other.
- Williamson (1991): This paper suggests the possibility that the dynamic capabilities and resource-based perspectives will play out in combination. The paper argues that in the long run, the best strategy for firms is to organize and operate efficiently.
- Leonard-Barton (1992): This paper considers core organizational capabilities in terms of: (1) employee knowledge and skills; (2) technical systems; (3) managerial systems, and (4) values and norms. The paper maintains that managers of new product and process development projects should take advantage of core capabilities while mitigating *core rigidities*. Twenty cases studies of new product and process development projects in five firms (e.g., Chaparral Steel, Ford Motor Company, and Hewlett Packard) provide illustrative data. [For students who find this topic of interest, Leonard-Barton (1995) is an exemplar research book.]
- Mahoney (1992a): This paper argues for an integrated organizational economic approach to strategic management based upon: (1) the behavioral theory of the firm; (2) transaction costs theory; (3) property rights theory; (4) agency theory; (5) resource-based theory/dynamic capabilities. Essentially this paper outlines the structure of the current book.
- Mahoney and Pandian (1992): Following Rumelt (1984), this paper argues that absent government intervention, isolating mechanisms (e.g., resource position barriers, invisible assets) exist because of asset specificity and bounded rationality.
- Amit and Schoemaker (1993): This paper adds behavioral decision-making biases and organizational implementation aspects as further impediments to the transferability or imitability of a firm's resources and capabilities.
- Peteraf (1993): This paper elucidates the organizational economics logic that is the foundation for the resource-based theory of Ricardian rents (Ricardo, 1817) and sustainable competitive advantage. The essence of the framework developed here is that four conditions must be met for achieving sustainable competitive advantage: (1) superior resources (firm heterogeneity within an industry), (2) *ex post* limits to competition (i.e., isolating mechanisms), (3) imperfect resource mobility (e.g., non-tradeable assets and co-specialized assets), and (4) *ex ante* limits to competition.
- Chi (1994): This paper develops a theoretical framework for analyzing the exchange structure in the trading of imperfectly imitable and imperfectly mobile firm resources. The paper first explores the conditions for such resources to be gainfully traded between

firms and then investigates the inter-connections between barriers to imitation and impediments to trading. A major part of the paper is devoted to developing a parsimonious and yet integrative (agency, property rights and transaction costs) model for assessing the exchange structure between firms that are involved in the trading of strategic resources in the face of significant transaction cost problems (such as adverse selection, moral hazard, contractual cheating, and hold-up problems that are due to information asymmetry, imperfect measurement, imperfect enforcement, and resource interdependencies).

- Farjoun (1994): This paper provides empirical support that unused productive services derived from human capital drives the diversification process. Unused productive services from existing human resources presents a “jig-saw puzzle” for balancing processes.
- Henderson and Cockburn (1994): Using both qualitative and quantitative data drawn from both public sources and from the internal records of 10 major European and American pharmaceutical firms, this paper attempts to measure the importance of heterogeneous, organizational capabilities. “Component” and “architectural” capabilities together explain a significant fraction of the variance in research productivity across firms.
- Godfrey and Hill (1995): This paper persuasively espouses the realist philosophy of science, which states that we cannot reject theories just because they contain key constructs that are unobservable.<sup>12</sup> It is not enough to state that the unobservability of utility dooms agency theory, that transaction costs theory is untestable because some transaction costs cannot be measured, or that resource-based theory is invalid because key resources (e.g., invisible assets) are unobservable. To reject a theory one must be able to show that the *predictions* of observable phenomena that are derived from the theory do not hold up under empirical testing.
- Mahoney (1995): This paper argues that the resource-based approach of deductive economics, the dynamic capabilities approach of strategy process, and organization

---

<sup>12</sup> In addition to Godfrey and Hill’s (1995) lucid discussion on realist philosophy, there are a number of works that cover various issues in philosophy of science and research methodology that are relevant to strategic management research and include: Blaug (1980), Caldwell (1984), Camerer (1985), Evered and Louis (1981), Huff (1981, 2000), Kaplan (1964), Kuhn (1970), Ladd (1987), Machlup, (1967), MacKinlay (1997), Mahoney (1993), Mahoney and Sanchez (1997, 2004), McCloskey (1993, 1998), McCloskey and Ziliak (1996), Montgomery, Wernerfelt and Balakrishnan (1989), Seth and Zinkhan (1991), Redman (1993), and Whetten (1989).



theory research on organizational learning (e.g., Argyris and Schon, 1978; Fiol and Lyles, 1985) need to be joined in the next generation of resource-based research.

- Zander and Kogut (1995): Based on their developed questionnaire distributed to project engineers knowledgeable of the history of 44 major innovations in 20 firms, this paper concludes that the transfer of manufacturing capabilities is influenced by the degree to which capabilities may be codified and taught. Empirical evidence corroborates the view that the nature of dynamic capabilities and the nature of competitive positioning matter.
- Foss (1996): This paper argues that there are complementarities between a contractual approach (e.g., transaction costs theory and property rights theory) and a knowledge-based approach (e.g., resource-based theory and knowledge-based theory) to strategic management. These complementarities are argued to be particularly fruitful for analyzing the strategic issues of the boundary and internal organization of the firm.
- Grant (1996): This paper argues that organizational capabilities are the outcome of knowledge integration: complex, team-based productive activities that cohesively integrate the knowledge of many individual specialists. Research in cross-functional capabilities in the context of new product development (Clark and Fujimoto, 1991) would be an exemplar.
- Miller and Shamsie (1996): This paper empirically tests resource-based theory in the context of the seven major United States film studios (i.e., MGM, Twentieth Century – Fox, Warner Brothers, Paramount, United Artists, Universal, and Columbia) from 1936 through 1965. This paper finds that property-based resources in the form of exclusive long-term contracts with stars and theaters helped financial performance in the stable environment of 1936-1950. In contrast, knowledge-based resources in the form of production and coordination talent boosted financial performance in the more uncertain post-television environment.
- Mowery, Oxley and Silverman (1996): Examining cross-citation rates for 792 partners in bilateral alliances that involved at least one U.S. firm and were established during 1985 and 1986, this paper provides empirical support for the importance of gaining capabilities through alliances. The empirical results bolster the argument that experience in related technological areas is an important determinant of absorptive capacity.
- Spender (1996): Building on Nelson and Winter (1982), and Nonaka and Takeuchi (1995), this paper views the firm as a dynamic knowledge-based activity system. The paper's arguments are consistent with Penrose's (1959) view of knowledge as the skilled process of leveraging resources, where that knowledge is embedded in the organization.
- Szulanski (1996): Based on 271 observations of 122 best-practice transfers in eight companies, the major barriers to internal knowledge transfer are found to be knowledge-

related factors such as the recipient's lack of absorptive capacity, causal ambiguity, and an arduous relationship between the source and the recipient.

- Helfat (1997): This empirical investigation of dynamic R&D capabilities examines the role of complementary know-how and other resources in the context of changing conditions in the U.S. petroleum industry during the 1970s and early 1980s. The empirical analysis indicates that in response to rising oil prices, firms with larger amounts of complementary technological knowledge and physical resources also undertook larger amounts of R&D on coal conversion (a synthetic fuel process).
- Powell and Dent-Micallef (1997): This paper examines the information technology literature, develops an integrative resource-based theoretical framework, and presents results from an empirical study of the retail industry. The empirical results support the view that information technology creates economic value by leveraging and utilizing complementary human and physical resources.
- Teece, Pisano and Shuen (1997): This paper views the dynamic capabilities perspective as building on Schumpeter (1934, 1950), Nelson and Winter (1982), and Teece (1982). Focal concerns are resource accumulation, replicability, and inimitability of organizational capabilities.
- Tripsas (1997): This paper analyzes the technological and competitive history of the global typesetter industry from 1886 to 1990. Key success factors include investment, technical capabilities, and appropriability through specialized complementary assets.
- Bogner, Mahoney and Thomas (1998): This paper, following Machlup (1967), argues that resource-based theory needs to move beyond: (1) theoretical construction that abstracts from historical time; (2) theory that focuses only on the stationary state; (3) theory where taxonomic and tautological arguments are made; (4) theory that focuses exclusively on the conditions for establishing equilibrium; and (5) theory that omits time as an independent variable.
- Farjoun (1998): This paper examines empirically the joint effect of skill-based and physical-based related diversification on accounting and financial measures of performance. For a sample of 158 large diversified manufacturing firms the joint effort of skill-based and physical-based related diversification had a strong positive effect on most indicators of performance. This finding corroborates resource-based theory that related diversification that builds on both skill-based and physical-based resources allows firms to create economic value by sharing and transferring these resources and to utilize activities and routines in which these resources interact.

- Lieberman and Montgomery (1998): Building on Lieberman (1987) and Lieberman and Montgomery (1988), this paper argues that resource-based theory and first-mover (dis)advantage are related conceptual frameworks that can benefit from closer linkages.
- Ahuja and Katila (1999): Using a sample of acquisition and patent activities of 72 leading firms from the global chemicals industry from 1980 to 1991, the relatedness of acquired and acquiring knowledge-based resources has a non-linear impact on innovation output. In particular, acquisition of firms with high levels of relatedness and un-relatedness both prove inferior to acquiring firms with moderate levels of knowledge-based relatedness.
- Argote (1999): This book presents evidence that organizations vary tremendously in the rate at which they learn. Argote (1999) argues that differences in patterns of knowledge creation, retention, and transfer contribute to differences in the rates at which organizations learn.
- Brush and Artz (1999): Using a sample of 193 veterinary practices, this paper investigates contingencies among resources, capabilities and performance in veterinary medicine. Empirical evidence supports the view that the economic value of resources and capabilities depends upon the information asymmetry characteristics of the product market.
- Silverman (1999): This paper considers how a firm's resource base affects the choice of industries into which the firm diversifies. The paper offers two main extensions of prior resource-based research. First, the paper operationalizes technological resources at a more fine-grained level than in prior empirical studies, thereby enabling a more detailed analysis concerning the *direction* of diversification. This analysis indicates that the predictive power of resource-based theory is greatly improved when resources are measured at a more fine-grained level. Second, the paper integrates transaction costs theory and resource-based theory to provide more detailed predictions concerning diversification. Empirical evidence suggests circumstances where resources (that have high asset specificity) can be and are utilized through contracting rather than through becoming a diversified firm.
- Williamson (1999): This paper suggests that one way of looking at research opportunities in strategic management is to view transaction costs theory as feeding into the organizational capabilities perspective. Both transaction costs theory and resource-based theory are viewed as needed in our efforts to understand complex business phenomena as we build towards a science of organization.
- Yeoh and Roth (1999): This paper empirically examines the impact of firm resources and capabilities using a sample of 20 pharmaceutical firms that operated as separate entrepreneurs between 1971 and 1989. The empirical results indicate that R&D and sales force expenditures have direct and indirect effects on sustainable competitive advantage.

- Bowman and Helfat (2001): This paper examines the resource-based theory that there is a significant role for corporate strategy based on the utilization of common resources by related businesses within a firm (Peteraf, 1993; Teece, 1982). Based on an analysis of the variance decomposition research literature, Bowman and Helfat (2001) conclude that corporate strategy (Andrews, 1980; Ansoff, 1965), in fact, does matter for economic performance.
- Makadok (2001): This paper provides a mathematical model synthesizing resource-based and dynamic capabilities views of economic value creation. Resource-picking (emphasized by resource-based theory) and capability-building (emphasized by the dynamic capabilities approach) for the purpose of achieving economic rent creation are shown to be complementary in some business circumstances but are shown to be substitutes in other business circumstances.
- Mahoney (2001): This paper argues that resource-based theory is primarily a theory of economic rents while transaction costs theory is primarily a theory of the existence of the firm. These two theories are complementary and are connected in the following way: resource-based theory seeks to delineate the set of market frictions that would lead to firm growth and sustainable economic rents (via isolating mechanisms), while transaction costs theory seeks to delineate the set of market frictions that explain the existence of the firm. The paper submits that the set of market frictions that explain sustainable firm rents (in resource-based theory) will be sufficient market frictions to explain the existence of the firm (in transaction costs theory). The paper also argues that the resource-based theory of the strategic (rent-generating and rent-sustaining) firm cannot assume away *opportunism*.
- Afuah (2002): This paper provides a model for mapping firm capabilities into competitive advantage. Using a sample of 78 observations for cholesterol drugs in the market from 1988 to 1994, this paper illustrates how the model can be used to estimate competitive advantage from technological capabilities.
- Coff (2002): Empirical results from a sample of 324 acquisitions that closed or failed to close in the years 1988 and 1989 offer evidence in support of the hypothesis that related human capital expertise between the acquirer and acquired enterprise can mitigate opportunism hazards associated with human capital asset specificity (Becker, 1964). In this business setting, related knowledge-based resources, in the form of related human expertise, increases the probability that a given transaction will close.
- Madhok (2002): This paper maintains that a strategic theory of the firm should address not just the decision with respect to hierarchical governance or market governance, but should also take into account how a firm's resources and capabilities can best be developed and deployed in the search for competitive advantage. Or put differently, transaction costs theory should be coupled with resource-based theory.

- Thomke and Kuemmerle (2002): Using a combination of field research, discovery data from nine pharmaceutical firms, and data on 218 alliances involving new technologies for experimentation and testing, several causes affecting resource accumulation are identified and described. The paper provides empirical support that the difficulty of imitating a particular resource is affected by the interdependencies with other resources.
- Adner and Helfat (2003): This paper adds to the study of competitive heterogeneity by measuring the economic effect of specific corporate-level managerial decisions, driven by dynamic managerial capabilities, on the variance of economic performance among U.S. energy companies. The empirical results also strongly suggest that corporate managers matter.
- Helfat and Peteraf (2003): This paper introduces the capability lifecycle, which identifies general patterns and paths in the evolution of organizational capabilities over time. The framework is intended to provide a theoretical structure for a more comprehensive approach to dynamic resource-based theory.
- Hoopes, Madsen and Walker (2003): This paper maintains that the resource-based view's accomplishments are clearer when seen as part of a larger theory of competitive heterogeneity. Combining economics, organization theory, and traditional business policy, the RBV suggests how, in a competitive environment, firms maintain unique and sustainable positions.
- Knott (2003a): This paper outlines a theory of sustainable innovation fueled by persistent heterogeneity. Knott (2003) shows that there exist conditions that generate persistent heterogeneity and sustainable innovation with each firm behaving optimally taking other firms' behaviors into account.
- Knott (2003b): This paper finds that franchising routines are both valuable and can lead to sustainable competitive advantage. The upshot of this empirical paper is that tacit knowledge is not necessary for having an isolating mechanism.
- Lippman and Rumelt (2003): This paper critiques the micro-foundations of neoclassical theory and develops further the concept of "rent." The paper also provides insights on rent sensitivity analysis, and a payments perspective of strategic management.
- Makadok (2003): This paper models mathematically the joint impact of two determinants of profitable resource advantages: the accuracy of managers' expectations about the future economic value of a resource, and the severity of agency problems that cause manager's interests to diverge from those of shareholders. The conclusion is that future research on the origins of competitive advantage should examine agency and governance issues along with, not apart from, resource-based issues.

- Szulanski (2003): This research book on *sticky knowledge* addresses an important question for managers: why don't best practices spread within organizations? Szulanski (2003) explores the effect of knowledge barriers and motivational barriers on knowledge transfer. Szulanski (2003) presents the empirical results of statistical analyses that stem from data collected through a two-step questionnaire survey. The research relies on 271 surveys studying the transfer of 38 (technical and administrative) practices in 8 companies. Szulanski (2003) finds that knowledge barriers to transfer have a larger effect on the stickiness of knowledge than motivational barriers, and the two barriers jointly explain nearly 75% of the variance in stickiness.

To conclude this chapter, we focus on a particularly important set of dynamic capabilities that are embedded in real options in strategic decision-making. Trigeorgis (1997) provides both rigor and relevance concerning strategic (real) options.

Trigeorgis, Lenos (1997). Real Options: Managerial Flexibility and Strategy in Resource Allocation. Cambridge, MA: MIT Press.

“Financial theory, properly applied, is critical to managing in an increasingly complex and risky business climate. ... Option analysis provides a more flexible approach to valuing our research investments. ... To me all kinds of business decisions are options.”

Judy Lewent, Chief Financial Officer, Merck & Co.  
Harvard Business Review (January-February, 1994)

Trigeorgis (1997) deals with the classical subject of resource allocation or project appraisal under uncertainty, particularly with the economic valuation of managerial operating flexibility and strategic actions as corporate *real options*. Similar to options on financial securities, real options involve “property rights” (with no obligations) to acquire or exchange an asset for a specified alternative price. The ability to value options (e.g., to defer, abandon, and grow) has brought a revolution to modern corporate finance theory on resource allocation.

Corporate value creation and competitive positioning are critically determined by corporate resource allocation, and by the proper evaluation of investment alternatives. Trigeorgis (1997) argues that traditional quantitative techniques such as discounted-cash-flow (DCF) analysis (that consider the size, timing and uncertainty of cash flows) have failed in business practice because these techniques traditionally have not properly captured managerial flexibility to adapt and revise later decisions in response to unexpected market developments. Moreover, these techniques traditionally neither capture the strategic value resulting from proving a technology viable nor capture the economic impact of project interdependencies and competitive interactions. In the Nelson and Winter (1982) sense, *organizational capabilities* that enhance adaptability and strategic positioning provide the infrastructure for the creation, preservation, and exercise of corporate real options that can have significant economic value.

Trigeorgis (1997) notes that, in practice, managers have often been willing to overrule traditional investment criteria in order to accommodate operating flexibility and other strategic decisions that managers consider just as valuable as direct cash flows. It is now widely recognized, for example, that traditional discounted-cash-flow (DCF) approaches to the appraisal of capital-investment projects, such as the standard net-present-value (NPV) rule of accepting positive NPV projects, does not properly capture management's flexibility to adapt and revise later decisions in response to unexpected market developments. Or put differently, a *theoretically accurate* NPV analysis would include real options values.

Trigeorgis (1997) argues that in the business marketplace, which is characterized by change, uncertainty, and competitive interactions, the realization of cash flows will probably differ from what management expected at the outset. As new information arrives and uncertainties about market conditions and future cash flows are gradually resolved, management may have valuable flexibility to alter its initial operating plan in order to capitalize on favorable future opportunities or to react so as to mitigate economic losses. For example, management may be able to defer, expand, contract, abandon, or otherwise alter a project at various stages of the project's useful operating life.

This managerial operating flexibility is likened to financial options. An American *call option* of an asset (with current value  $V$ ) gives the right, with no obligation, to acquire the underlying asset by paying a pre-specified price (the *exercise price*,  $I$ ) on or before a given maturity. Similarly, an American *put option* gives the right to sell (or exchange) the underlying asset and receive the exercise price. *The asymmetry deriving from having the right but not the obligation to exercise an option is at the heart of the option's value.*



Trigeorgis (1997) notes that as with options on financial securities, management's flexibility to adapt its future actions in response to altered future market conditions and competitive reactions expands a capital-investment opportunity's value by improving its upside potential while limiting its downside economic losses relative to the initial expectations of a passive management. The resulting asymmetry calls for a *strategic investment criterion*, reflecting both value components: (1) the traditional static NPV of direct cash flows, and (2) the real option value of operating flexibility and strategic interactions.

Trigeorgis (1997) argues that a real options approach to capital budgeting has the potential to conceptualize and quantify the value of options from active management and strategic interactions. This economic value is typically manifest as a collection of "real options" embedded in capital-investment opportunities, having as the underlying asset the gross project value of discounted expected operating cash inflows. Many of these real options (e.g., to defer, contract, shut down, or abandon a capital investment) occur naturally; other real options may be planned and built in at some extra cost from the outset (e.g., to expand capacity, to build growth options, to default when investment is staged sequentially, or to switch between alternative inputs or outputs.). Let us now consider various real options.

1. *Option to Defer Investment.* The real option to defer an investment decision is analogous to an American call option on the gross present value of the completed project's expected operating cash flows,  $V$ , with an exercise price equal to the required outlay,  $I$ . Management holds a lease on (or an option to buy) valuable land or resources. Management can wait  $x$  years to see if output prices justify constructing a building or a plant or developing a field. The option to wait is particularly valuable in natural-resource extraction industries, farming, paper products, and real estate development.

Consider the following example from Dixit and Pindyck (1994): How should we decide whether or not to enter into a business? If we refer to the literature on finance, the traditional approach is to use cash flow analysis using a net present value criterion. For example, let us imagine a situation in which we are considering entering the business of making "widgets." Assume that it costs \$1600 to build a widget factory and that our current cost of capital is 10 percent. In addition, we sell only one widget per year, and the current price of a widget is \$200. While we know the current price for widgets, we are somewhat uncertain about the future prices. Forecasts indicate that there is a 50% chance that prices will go up to \$300 next period (and remain there forever), however, there is also a 50% chance that prices will go down to \$100. This forecast implies that the expected price of widgets in the future is \$200 ( $= .5 * \$300 + .5 * \$100$ ).

Using these numbers, we can evaluate this "project" with a standard cash flow analysis. The expected cash flow from entering the widget business appears in the first column of Table 1. In period 0, we build the plant (-\$1,600) and begin production, receiving \$200 in revenues ( $-\$1,600 + \$200 = -1,400$ ). From that period on, we have positive *expected cash flow* of \$200. We can use this cash flow series to arrive at the net present value (NPV) for the project, which is \$600. (Since the value at  $T_0$  of a perpetuity CF beginning at  $T_1$  with a discount rate  $r$  equals  $CF/r$  then here it is  $\$200/.1 = \$2000$ ; then we take this \$2,000 and subtract \$1,400 to arrive at \$600.) We would then proceed with the project since the NPV of \$600 is greater than zero.

However, what if we wait a period, and find out whether the price goes up or down? That is, what if we choose to keep our options open and remain flexible in our decision. Two different scenarios could occur. The first possibility is that the price goes up to \$300, in which case we would experience the cash flow under Scenario 1 in Table 1. The second possibility is that the

price goes down to \$100; in which case, we obtain the cash flows under scenario 2. Now, one will notice that under scenario 1, the NPV (in period 0) is positive (i.e.,  $NPV = \$1,545$ ); however, under scenario 2, the NPV is negative (i.e., NPV is  $-\$455$ ). (The present value of the perpetuity is  $\$300/.1 = \$3,000$  from which we subtract  $\$1,600/1.1 = -\$1,455$  to arrive at  $\$1,545$  and for the low demand scenario we have  $\$100/.1 = \$1,000$  and subtract  $\$1,600/1.1 = -\$455$ ). Thus, if we waited a period and the price went up to \$300, we would proceed with the project; while if the price went down, we would not proceed with the project. Thus, under the second scenario, the actual NPV would *not* be  $-\$455$ , but would be  $\$0$ ; that is, we would not invest in a negative NPV project. What does this tell us about the value of waiting, and remaining strategically flexible?

One way of answering this question is to reframe our cash flow analysis. Instead of taking the *NPV of the expected cash flows*, let us calculate the *expected NPV* of the two scenarios combined. That is, we have a 50% chance of the price going up and getting an NPV of  $\$1,545$ , and a 50% chance of the price going down and getting  $\$0$ . The expected combined NPV is therefore approximately  $\$773 (= .5 \times \$1,545 + .5 \times 0)$ . The NPV where we wait, find out the true price, and then make the decision is larger (by  $\$173$ ) than going right ahead right now. There is (an option) value to waiting of  $\$173$ . Thus, we can increase our expected returns by waiting a year and then deciding whether to undertake the sunk-cost investments in a new plant.

**Table 1**

<b>Time</b>	<b>Expected Cash Flow (Traditional NPV)</b>	<b>Expected Cash Flow (Scenario 1)</b>	<b>Expected Cash Flow (Scenario 2)</b>
0	\$ (1,400)	\$ --	\$ --
1	\$ 200	\$ (1,300)	\$ (1,500)
2	\$ 200	\$ 300	\$ 100
3	\$ 200	\$ 300	\$ 100
4	\$ 200	\$ 300	\$ 100
5	\$ 200	\$ 300	\$ 100
6	\$ 200	\$ 300	\$ 100
7	\$ 200	\$ 300	\$ 100
8	\$ 200	\$ 300	\$ 100
Inf	\$ 200	\$ 300	\$ 100
<b>NPV</b>	<b>\$ 600</b>	<b>\$ 1,545</b>	<b>\$ (455)</b>

*Summary ideas on this first option of waiting (and learning) before investing:* The above example illustrates that even when the static (positive) NPV calculation suggests a "Go," when the real options value of flexibility is taken into account, the top-level manager should wait. The option to wait is equivalent to a call option on the investment project. The call is exercised when the firm commits to the project. But often it is better to defer a positive-NPV project in order to keep the call option alive. Deferral is most attractive when uncertainty is great and immediate project cash flows -- which are lost or postponed by waiting -- are small.

2. *Option to Default during Staged Construction (Time-to-Build Option).* Each stage of an investment can be viewed as an option on the economic value of subsequent stages by incurring the installment-cost outlay (e.g.,  $I_1$ ) required to proceed to the next stage, and can therefore be valued similar to options on options (or compound options). Staging the investment as a series of outlays creates the real option to abandon the project in mid-stream if new

information is unfavorable. This real option is valuable in R&D-intensive industries (especially pharmaceuticals), in highly uncertain long-development, capital-intensive industries (such as energy-generating plants or large-scale construction), and in venture capital.

3. *Option to Expand, to Contract to Shut Down and Restart Operations.* If market conditions are more favorable than expected, the firm can expand the scale of production or accelerate resource utilization. Conversely, if conditions are less favorable than expected, the firm can reduce the scale of operations. In extreme cases, production may be halted and restarted. Applications can be found in: natural-resource industries (e.g., mining); facilities planning and construction in cyclical industries; fashion apparel; consumer goods; and commercial real estate.

4. *Option to Abandon for Salvage Value.* Management may have a valuable real option to abandon a project in exchange for its salvage value. Naturally, more general-purpose capital assets would have a higher salvage value and abandonment option value than special-purpose assets. Valuable abandonment options are generally found in capital-intensive industries (such as airlines and railroads), in financial services, and in new-product introductions in uncertain markets. Abandonment should not be exercised lightly if abandonment might lead to eventual erosion of valuable expertise and other crucial organizational capabilities that could be applied elsewhere in the business or that could prevent the firm from participating in future technological developments. Moreover, abandonment may lead to the loss of goodwill from customers.

5. *Option to Switch Use (e.g., Inputs or Outputs).* Generally, process flexibility can be achieved not only via technology (e.g., by building a flexible facility that can switch among alternative energy inputs) but also by maintaining relationships with a variety of suppliers and switching among them as their relative prices change. Process flexibility is valuable in feedstock-dependent facilities, such as oil, electric power, chemicals, and crop switching. Product

flexibility, enabling the firm to switch among alternative outputs, is more valuable in industries such as machine parts, automobiles, consumer electronics, toys, specialty paper, and pharmaceuticals, where product differentiation and diversity are important and/or product demand is volatile. In such business cases it might be worthwhile to install a more costly flexible capacity to acquire the dynamic capability to alter product mix or production scale in response to changing market conditions.

6. *Corporate Growth Options.* Corporate growth options that set the path of future opportunities are of considerable strategic importance. Although in isolation a proposed facility may appear unattractive, such a facility may be only the first in a series of similar facilities if the process is successfully developed and commercialized, and it may even lead to entirely new by-products. Many early investments (e.g., in R&D) can be seen as prerequisites or links in a chain of interrelated projects. The value of the early projects derives not so much from their expected directly measurable cash flows as from the future growth opportunities they may unlock (e.g., access to a new market, or strengthening of the firm's core capabilities and its strategic positioning). An opportunity to invest in a first-generation high-tech product, for example, is analogous to an option on options (an inter-project compound option). Despite a negative static NPV, the infrastructure, experience, and potential by-products generated during the development of the first-generation product may serve as springboards for developing lower-cost or higher-quality future generations of that product, or even for generating entirely new applications. But unless the firm makes that initial investment, subsequent generations or other business applications will not even be feasible. The infrastructure and experience gained, if proprietary can place the firm at a competitive advantage, which may even reinforce itself if learning-cost-

curve effects are present. Growth options are found in all infrastructure-based or strategic industries --- especially in high technology, in R&D, in industries with multiple product generations or applications (e.g., semi-conductors, computers, pharmaceuticals), in multinational operations, and in strategic acquisitions.

7. *Multiple Interacting Options.* Real-life projects often involve a collection of various options. Upward potential-enhancing and downward-protection options are present in combination. Their combined economic value may differ from the sum of their separate values, i.e., they interact. They may also interact with financial flexibility options. Applications include most industries listed above.

Trigeorgis (1997) argues that real options have the potential to make a significant difference in strategic management. Sustainable competitive advantages resulting from proprietary technologies, scale, ownership of valuable natural resources, managerial capital, reputation, brand name, or patents (Andersen, 2001; Arora, Fosfuri and Gambardella, 2001) empower companies with valuable real options to grow through future profitable investments and to more effectively respond to unexpected adversities or opportunities in a changing technological, competitive, or general business environment. Students studying the economics of organization have ample opportunity to supplement real options analysis (that is often decision-theoretic) with game-theoretic tools capable of incorporating strategic competitive responses, and this research area promises to be an important and challenging direction for strategic management and corporate finance research.

---

### **Applications of the Real Options Perspective:**

- In 1984 the W. R. Grace Corporation made an investment in a new technology for automotive catalytic converters. Although the technology proved uncompetitive on price in the automotive market, new applications arose in co-generation plants and in utility emission control as a result of the U.S. Clean Air Act.
- In research and development, many high-technology companies invest heavily in technologies that may result in a wide range of possible outcomes and new potential markets, but with a high probability of technical or market failure. In the pharmaceutical industry, for example, on average it costs \$360 million and takes a decade to bring a new drug to the market. Only one explored chemical in 10,000 becomes a prescription drug, and once a drug reaches the market the drug faces a 70% chance of failing to earn the cost of invested capital. Such investments are hard to sell to top management on financial grounds; their benefits are remote and hard to quantify, even though intuitively their growth potential seems promising. Instead of ignoring these technologies, a company can make a capital commitment in stages, effectively taking a call option on the underlying technology (or future applications). The initial outlay is not made so much for its own cash flows as for its growth-option value.
- Merck and Company embarked on extensive automation, starting with a drug packaging and distribution plant, even though technical success was uncertain and projected labor savings did not seem to justify the investment. Operations valuation allowed engineers to articulate the whole range of outcomes and their benefits. Indeed, building and using real options-based planning methods were viewed as having created a valuable new capability for Merck (Nichols, 1994). In fact, the more uncertain the technology or the future market demand, the higher the value of the real option.



- **Nucor.** The case of the adoption of the thin-slab caster by Nucor involved consideration of sunk-cost commitments and real options (Ghemawat, 1997):
  - **The option to wait (and learn) before investing.** In the Nucor case, it was very unlikely that another firm would be willing to be the pioneering firm to deploy this new technology. Thus, if Nucor were to wait, the reduction in uncertainty would have been small.
  - **The option to make follow-on investments if the immediate investment project succeeds.** In the Nucor case, first plant appeared to have a slightly negative NPV on a stand-alone basis across a majority of likely scenarios. However, the experience gained in building the first plant would substantially improve the economics of subsequent plants. Thus, the first plant could merely be the "price of admission" representing a necessary learning curve. Thus, even though the narrow (negative) NPV calculation for Nucor suggested a "NO GO" the **growth options** tipped the scale to "GO." Nucor took into account the "strategic" value of taking on this negative-NPV project. A close look at Nucor's payoffs reveals a call option on follow-on projects in addition to the immediate project's cash flows. Today's investments can generate tomorrow's opportunities.
  - **The option to abandon the project.** Even if the CSP fails, that component of the mill could be potentially replaced by another technology; the bulk of the mill, such as electric arc furnace and rolling mills, may be useable even with another thin-slab technology. Thus, when the narrow (negative) NPV calculation suggests a "NO GO" a high options value of abandonment (i.e., low switching costs) may tip the scale to "GO." The option to abandon a project provides partial insurance against failure. This is a put option; the put option's exercise price is the value of the project's assets if sold or shifted to a more valuable use.

The flexibility provided by flexible manufacturing systems, flexible production technology, or other machinery having multiple uses can be analyzed from the real options perspective. Recently, the flexibility created by modular design that connects components of a larger system through standardized interfaces (and its impact on organization design) has captured attention in strategic management (Baldwin and Clark, 2000; Bowman and Kogut, 1995; Garud and Kumaraswamy, 1995; Garud, Kumaraswamy, and Langlois, 2003; Langlois, 2002; Sanchez and Mahoney, 1996, 2001; Schilling, 2002; Worren, Moore and Cardona, 2002).

Students studying the economics of organization have an opportunity to evaluate such flexibility using the real options framework.

In conclusion, the current academic research literature in corporate finance has largely framed real-options problems as *decision theoretic*. However, we now need to move on to considerations where the timing of investments also depends on how other players will respond. Thus, strategic management must take into account both decision-theoretic problems and *game-theoretic* problems in the next generation of real options research.<sup>13</sup>

Concluding Comments. The resource-based, dynamic capabilities and real options literatures are potentially highly synergistic for theory development, empirical testing, and business applications. Students with research interests in business history (e.g., Chandler, 1990), evolutionary theory and organizational capabilities (e.g., Nelson and Winter, 1982), corporate finance (e.g., Trigeorgis, 1997), strategic human resource management (Baron and Kreps, 1999), and entrepreneurship (Penrose, 1959) are anticipated to contribute to the evolving science of organization.

---

<sup>13</sup> For strategic management contributions to the real options perspective see: Bowman and Hurry (1993); Chi (2000); Folta (1998); Folta and Miller (2002); Kogut (1991); McGrath (1997, 1999); Miller (2002), Miller and Folta (2002), and Sanchez (1993, 2003). Schwartz and Trigeorgis (2001) provide a number of classical readings and recent contributions on real options, and investment under uncertainty.

In this research book, I have maintained that an integrated organizational economics research program is a feasible, challenging, and rewarding endeavor in pursuing the evolving science of organization. Toward that end, five major theories of the firm are reviewed and interrelated: (1) the behavioral theory of the firm; (2) transaction costs theory; (3) property rights theory; (4) agency theory; and (5) dynamic resource-based theory. In my judgment, content research (e.g., deductive agency and transaction costs economics) and process research (e.g., research on cognitive psychology and the behavioral theory of the firm) need to be joined in the next generation of organizational economics research.

My major philosophical position in this research book is that the burden of proof for claiming that theories are incommensurable (Kuhn, 1970) should be placed on those scholars who make such claims. Thus, I concur with Popper's (1970) strong dissent of Kuhn's (1970) incommensurability thesis. In my judgment, bridges can and should be built between contested intellectual terrains.

While this research book provides a rudimentary first step toward intertwining the various strands of the organizational economics research literature into one cord, I anticipate that the next generation of students will go well beyond the mere comprehension of the various theories of the firm and will successfully apply the organizational economics perspective to emerging business phenomena (e.g., in research studies of information technology and its impact on organizational structures and to the study of transition economies).

In fact, the growing interest in the organizational economics view of strategic management is arguably due to the fruitfulness of organizational economics theory in

contributing to our understanding of currently observed changes at the organizational and institutional levels of analysis. The reason for such a knowledge claim is not hard to find. Even though technological, organizational, and institutional change advances breathlessly, the organizational economics principles provided in this research book are durable principles that have stood, and will continue to stand, the test of time. Moreover, I also anticipate that the next generation of students pursuing the evolving science of organization will be largely successful in supplying a more unified organizational economics approach to strategic management.

If the benefits of combining content and process are so transparently “obvious,” as I claim that they are in this concluding chapter, then one might reasonably ask: why didn’t industrial organization economics take this path toward combining content and process a long time ago? My own view is that much of the answer to this question resides in the historical fact that many of the early influential economists in academia had an expressed desire to build a deductive economic science that would be built from a few premises. Simon’s (1957: 198) comments are apposite:

The reluctance of economic theory to relinquish its classical model of economic man is understandable. When even a small concession has been made in the direction of admitting the fallibility of economic man, his psychological properties are no longer irrelevant. Deductive reasoning then no longer suffices for the unique prediction of his behavior without constant assistance from empirical observation.

Once the cognitive limitations of decision-makers are admitted, then developing a useful science of organization may begin.

In the behavioral theory of the firm (in chapter 1) we see that those researchers operating from the behavioral perspective tend to view the organization as a more efficient information processor than the individual. The firm, for example, is considered to be a functionally rational

institutional response to uncertainty and bounded rationality (Cyert and March, 1963; Simon and March, 1958). Indeed, Thompson notes that: “Uncertainty appears as the fundamental problem for complex organizations, and coping with uncertainty, (is) the essence of the administrative process (1967: 159). Organizations are structures of mutual expectation that reduce equivocality.

Transaction costs theory (in chapter 2) builds a theory of the firm based on a core premise of the behavioral theory of the firm --- bounded rationality. To make information flow considerations even more problematic, transaction costs theory holds that the human limitations for processing information are aggravated further by the potential for deliberate non-disclosure of information, deliberate obfuscation, and the making of self-disbelieved statements --- all of which are forms of opportunism. The firm is then viewed not only as a mechanism to take account of bounded rationality (in common with the behavioral theory of the firm), but also to attenuate opportunistic behavior. The firm may be preferred to the market due to the firm’s adaptability, monitoring, dispute settling, and reward refining attributes.

A theory related to the transaction costs theory is property rights theory (in chapter 3). Indeed, two of the more important historical figures in transaction costs theory have seminal works in property rights theory (Coase, 1960; Commons, 1924). In the property rights approach, the corporation is viewed as “a method of property tenure” (Berle and Means, 1932: 1).

The three criteria for efficiency of property rights are: (1) universality -- all scarce resources are owned by someone; (2) exclusivity – property rights are exclusive rights; and (3) transferability – to ensure that resources can be allocated from lower to higher yield uses. In neoclassical microeconomic theory, all three criteria are assumed to hold. However, in the real world of positive transaction costs, such as measurement costs (Barzel, 1989) none of these criteria are ensured.

Alchian and Demsetz (1972) argue that the structure of property rights has a significant impact on agency and transaction costs and that property rights establish an institutional context within which transactions are negotiated. Agency theory, transaction costs theory and property rights theory, thus, are highly intertwined. Indeed, the significance of the study of property rights theory results from the fact that positive transaction costs are present (Coase, 1960).

There are several commonalities between transaction costs theory and agency theory (in chapter 4) --- especially with *positive* agency theory (Jensen, 1983; Jensen and Meckling, 1977). Both theories assume bounded rationality and opportunism (e.g., moral hazard). The consequences of bounded rationality and opportunism in both theories are incomplete contracting and the potential for contractual hazards. Agency theory emphasizes ex ante incentive alignment, while transaction costs theory emphasizes ex post governance issues (e.g., dispute resolution and mal-adaptation costs).

Dynamic resource-based theory (chapter 5) is intimately tied to all four previous theories of the firm. Resource-based theory is linked to the behavioral theory of the firm, if superior heuristics lead to sustainable competitive advantage (Schoemaker, 1990). Moreover, the more evolutionary economics component of dynamic resource-based theory as developed by Nelson and Winter (1982) draws heavily from the Carnegie School behavioral theory of the firm (Cyert and March, 1963; March and Simon 1958) with its emphasis on “routines” and “search.” Resource-based theory is linked to transaction costs theory because resource combinations are influenced by transaction costs considerations (Teece, 1982). Resource-based theory is linked to property rights theory since well-delineated property rights make resources economically valuable, and as resources become more economically valuable, property rights become more

precise (Libecap, 1989). Finally, resource-based theory is linked to agency theory since economic incentives highly influence the deployment of resources and the development of organizational capabilities (Castanias and Helfat, 1991).

This concluding chapter has emphasized the complementarities among these organizational economics theories of the firm and the potential for integration in the evolving science of organization (Langlois, Yu and Robertson, 2002; Tan and Mahoney, 2003). I hasten to add, however, that each of these five theories has distinctive features. The behavioral theory focuses almost exclusively on the bounded rationality problem, while transaction costs theory focuses primarily on the attenuation of opportunism (albeit acknowledging the bounded rationality problem).

Property rights theory, like transaction costs theory, holds the premises of both bounded rationality and opportunism, but property rights theory is at a more aggregate unit of analysis. Property rights theory focuses more on the institutional level of analysis, while transaction costs theory focuses at a more micro-analytic level of firm governance and the transaction. This difference in unit of analysis helps explain why property rights theory (e.g., Libecap, 1989; North 1990) emphasizes inertia and the lack of adaptation, while transaction costs theory (Williamson, 1975, 1996) emphasizes more that firm-level capabilities adapt over time to achieve relatively superior economic efficiencies.

Agency theory, in the form of the *mathematical* principal-agent model, which was highlighted in chapter 4, assumes perfect rationality (i.e., no bounded rationality problems are assumed to exist). Clearly, such a conceptual lens may magnify and clarify economic incentives issues, but the theory will also inevitably blur some managerial problems, and indeed will even neglect managerial problems of information-processing limitations.

Finally, I have recently argued (Mahoney, 2001) that while resource-based theory holds the same premises as transaction costs theory --- bounded rationality and opportunism --- the theories are distinguished by the phenomena to be explained. Resource-based theory is a theory of firm-level economic rents, and transaction costs theory is a theory of the existence of the firm. I also make the conjecture that the set of market frictions to explain sustainable firm-level rents (in dynamic resource-based theory) will be sufficient market frictions to explain the existence of the firm (in transaction costs theory). Thus, these five theories of the firm are unique. The argument advanced here is that the five theories developed in this research book can be thought of as “modular building blocks” to develop new intellectual combinations of thought in the evolving science of organization.<sup>13</sup>

**Conclusions.** The study of the economics of organization offers a wide range of challenges for students. In my judgment, at a general level, content (e.g., deductive economics) and process (e.g., behavioral and cognitive) approaches need to be joined in the next generation of organizational economics research. Moreover, the economics of organization requires far greater attention to the interactions between the institutional environment (e.g., North, 1990) and organizational-level governance (e.g., Williamson, 1996).

---

<sup>13</sup> I thank Anne Huff for suggesting the idea that the structure of this research book can be usefully described as having a “modular architecture,” where doctoral students may use various combinations of these theories of the firm depending on the problem at hand. In addition, Anne also points out that it is worth noting that each of the five theories has its own precise vocabulary and specific phenomena of interest and thus doctoral students may continue to contribute analytically by specializing in one of these theories in addition to the possibilities of synthesis, which I have emphasized throughout this book.



We conclude with the central messages from the economics of organization, followed by some key questions for the strategy management field. The core ideas of the organizational economics approach are:

1. Behavioral Theory of the Firm: Get the Routines Right to Operate Effectively;
  2. Transaction Costs Theory: Get the Governance Right to Minimize Costs;
  3. Property Rights Theory: Get the Property Rights Well-Defined and Correctly Allocated to Create Wealth;
  4. Agency Theory: Get the Incentives Right to Minimize the Agency Loss;
  5. Resource-Based Theory: Choose the Right Resources (i.e., Valuable Rare, Inimitable, Non-Substitutable) to Generate and Sustain (Ricardian) Rents;
- Dynamic Capabilities Theory: Build Technological and Organizational Capabilities for (Schumpeterian) Rents; and
- Real Options Theory: Include Real Options Values to Analyze the Strategic Value of Investments Under Uncertainty.

We conclude with the following list of questions for further research using an organizational economics approach to strategic management:

- Routines: What are they? What role do routines play in organizational survival and sustainable competitive advantage? What are the refutable implications?
- Institutions: What are they? How do they differ? To what purpose and effect? Where does the action reside? What are the mechanisms? What are the refutable implications? What are the public-policy ramifications? What do the data support?"
- How can we describe, explain (and perhaps predict) how institutional (property rights) environments and governance (transaction costs and agency costs) play out?
- How does the interaction of managers and other human resources influence a firm's growth and competitive advantage?

- Why do firms differ in resources, capabilities, and economic performance?
- How can networks and strategic alliances be formed and organized to share and/or co-develop capabilities?
- What are the implications for firm growth of developing managers internally rather than hiring managers externally?
- To what extent can firms gain and sustain competitive advantage (via both competition and cooperation) by *strategic commitment*?
- How can firms gain and sustain competitive advantage by *strategic flexibility*? How can real options research help strategic management to place an economic value on these real options?
- How will the economic surplus generated by the firm be allocated among its members?
- If the many stakeholders who comprise the nexus of contracts are residual claimants (e.g., workers who have invested firm-specific human capital), then why are shareholders necessarily the ones affected the most by the firm's decisions? Even if they are, are these shareholders the party that benefits the most economically from the additional protection granted by the decision rights? (Zingales, 2000).

These are challenging questions for further inquiry. I anticipate that this generation of students studying the economics of organization will contribute greatly to the evolving science of organization. We can and will do better.

## References

- Adner, Ron and Constance E. Helfat (2003). Corporate effects and dynamic managerial capabilities. Strategic Management Journal, 24: 1011-1025.
- Afuah, Allan (2001). Dynamic boundaries of the firm: Are firms better off being vertically integrated in the face of a technological change? Academy of Management Journal, 44 (6): 1211-1228.
- Afuah, Allan (2002). Mapping technological capabilities into product markets and competitive advantage: The case of cholesterol drugs. Strategic Management Journal, 23: 171-179.
- Ahuja, Gautam and Ritta Katila (2001). Technological acquisitions and the innovation performance by acquiring firms: A longitudinal study. Strategic Management Journal, 22: 197-220.
- Akerlof, George A. (1970). The market for 'lemons': Qualitative uncertainty and the market mechanism. Quarterly Journal of Economics, 84: 488-500.
- Alchian, Armen A. (1950). Uncertainty, evolution, and economic theory. Journal of Political Economy, 58: 211-222.
- Alchian, Armen A. (1965). Some economics of property rights. Il Politico, 30: 816-829.
- Alchian, Armen A. and Harold Demsetz (1972). Production, information costs, and economic organization. American Economic Review, 62: 777-795.
- Alchian, Armen A. and Harold Demsetz (1973). The property right paradigm. Journal of Economic History, 33: 16-27.
- Allison, Graham T. (1971). Essence of Decision. Boston, MA: Little, Brown.
- Amit, Raphael and Paul J. H. Schoemaker (1993). Strategic assets and organizational rent. Strategic Management Journal, 14: 33-46.
- Anand, Jaideep and Andrew Delios (2002). Absolute and relative resources as determinants of international acquisitions. Strategic Management Journal, 23: 119-134.
- Anand, Jaideep and Harbir Singh (1997). Asset redeployment, acquisitions and corporate strategy in declining industries. Strategic Management Journal, 18 (Summer): 99-118.
- Andersen, Birgitte (2001). Technological Change and the Evolution of Corporate Innovation: The Structure of Patenting 1890-1990. Northampton, MA: Edward Elgar.
- Anderson, Erin (1985). The salesperson as outside agent or employee: A transaction cost analysis. Marketing Science, 4 (3): 234-254.
- Anderson, Erin and David Schmittlein (1984). Integration of the sales force: An empirical examination. Rand Journal of Economics, 15 (Autumn): 385-395.

- Andrews, Kenneth R. (1980). The Concept of Corporate Strategy. Homewood, IL: Irwin.
- Ansoff, H. Igor (1965). Corporate Strategy: An Analytical Approach to Business Policy for Growth and Expansion. New York, NY: McGraw-Hill Book Company.
- Argote, Linda (1999). Organizational Learning: Creating, Retaining and Transferring Knowledge. Boston: Kluwer.
- Argyres, Nicholas S. (1996). Evidence on the role of firm capabilities in vertical integration decisions. Strategic Management Journal, 17: 129-150.
- Argyres, Nicholas S. and Julia Porter Liebeskind (1999). Contractual commitments, bargaining power, and governance inseparability: Incorporating history into transaction cost theory. Academy of Management Review, 24 (1): 49-63.
- Argyris, Chris and Donald Schon (1978). Organizational Learning: A Theory of Action Perspective. Reading, MA: Addison-Wesley.
- Armour, Henry O., and David J. Teece (1978). Organizational structure and economic performance: A test of the multidivisional hypothesis. Bell Journal of Economics, 9: 106-122.
- Arora, Ashish, Andrea Fosfuri, and Alfonso Gambardella (2001). Markets for Technology: The Economics of Innovation and Corporate Strategy. Cambridge, MA: MIT Press.
- Arrow, Kenneth J. (1971). Essays in the Theory of Risk-Bearing. New York, NY: North-Holland Publishing Company.
- Arrow, Kenneth J. (1974). The Limits of Organization. New York, NY: W. W. Norton and Company.
- Arrow, Kenneth J. (1985). "The Economics of Agency" (pp.37-51). In Principals and Agents: The Structure of Business, edited by John W. Pratt and Richard J. Zeckhauser. Boston, MA: Harvard Business School Press.
- Arthur, W. Brian (1994). Increasing Returns and Path Dependence in the Economy. Ann Arbor, MI: University of Michigan Press.
- Bain, Joe S. (1968). Industrial Organization. New York, NY: John Wiley & Sons.
- Balakrishnan, Srinivasan and Birger Wernerfelt (1986). Technical change, competition and vertical integration. Strategic Management Journal, 7: 347-359.
- Baldwin, Carliss Y. and Kim B. Clark (2000). Design Rules: The Power of Modularity. Cambridge, MA: MIT Press.
- Barnard, Chester I. (1938). The Functions of the Executive. Cambridge, MA: Harvard University Press.

- Barnard, Chester I. (1948). Organization and Management: Selected Papers. Cambridge, MA: Harvard University Press.
- Barney, Jay B. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17 (1): 99-120.
- Barney, Jay and William G. Ouchi (Eds.) (1986). Organizational Economics: A New Paradigm for Understanding and Studying Organizations. San Francisco, CA: Jossey-Bass Publishers.
- Baron, James N. and David M. Kreps (1999). Strategic Human Resources: Frameworks for General Managers. New York, NY: John Wiley & Sons.
- Barzel, Yoram (1982). Measurement cost and the organization of markets. Journal of Law and Economics, 25 (April): 27-48.
- Barzel, Yoram (1989). Economic Analysis of Property Rights. Cambridge: Cambridge University Press.
- Baumol, William J., John C. Panzar and Robert D. Willig (1982). Contestable Markets and the Theory of Industrial Structure. New York, NY: Harcourt Brace Jovanovich.
- Becker, Gary S. (1964). Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education. New York, NY: Columbia University Press.
- Bensaou, Ben M. and Erin Anderson (1999). Buyer-supplier relations in industrial markets: When do buyers risk making idiosyncratic investments. Organization Science, 10 (4): 460-481.
- Berle, Adolf A. and Gardiner C. Means (1932). The Modern Corporation and Private Property. New York, NY: Macmillan.
- Blair, Roger D. and David L. Kaserman (1983). Law and Economics of Vertical Integration and Control. New York, NY: Academic Press.
- Blaug, Mark (1980). The Methodology of Economics. Cambridge: Cambridge University Press.
- Bogner, William C., Joseph T. Mahoney and Howard Thomas (1998). Paradigm shift: The parallel origin, evolution, and function of strategic group analysis with the resource-based theory of the firm. Advances in Strategic Management, 15: 63-102.
- Boulding, Kenneth E. (1950). A Reconstruction of Economics. New York, NY: John Wiley & Sons, Inc.
- Bowman, Edward H. and Constance E. Helfat (2001). Does corporate strategy matter? Strategic Management Journal, 22: 1-23.
- Bowman, Edward H. and Dileep Hurry (1993). Strategy through the options lens: An integrated view of resource investments and the incremental-choice process. Academy of Management Review, 18 (4): 760-782.
- Bowman, Edward H. and Bruce M. Kogut (Eds.) (1995). Redesigning the Firm. New York, NY: Oxford University Press.

- Brush, Thomas H. and Kendall W. Artz (1999). Toward a contingent resource-based theory: The impact of information asymmetry on the value of capabilities in veterinary medicine. Strategic Management Journal, 20: 223-250.
- Brynjolfsson, Erik (1994). Information assets, technology, and organization. Management Science, 40 (12): 1645-1662.
- Caldwell, Bruce (1984). Appraisal and Criticism in Economics. Boston, MA: Allen and Unwin.
- Camerer, Colin (1985). Redirecting research in business policy and strategy. Strategic Management Journal, 6: 1-15.
- Carney, Mick and Eric Gedajlovic (1991). Vertical integration in franchise systems: Agency theory and resource explanations. Strategic Management Journal 12: 607-629.
- Castanias, Richard P. and Connie E. Halfat (1991). Managerial resources and rents. Journal of Management, 17: 155-171.
- Caves, Richard E. and Ralph M. Bradburd (1988). The empirical determinants of vertical integration. Journal of Economic Behavior and Organization, 9: 265-279.
- Caves, Richard E. and William F. Murphy (1976). Franchising, firms, markets, and intangible assets. Southern Economic Journal, 42: 572-586.
- Chandler, Alfred D. (1962). Strategy and Structure. Cambridge, MA: MIT Press.
- Chandler, Alfred D. (1977). The Visible Hand: The Managerial Revolution in American Business. Cambridge, MA: Harvard University Press.
- Chandler, Alfred D. (1990). Scale and Scope: The Dynamics of Capitalism. Cambridge, MA: Harvard University Press.
- Chandler, Alfred D. (1992). Organizational capabilities and the economic history of the industrial enterprise. Journal of Economic Perspectives, 6 (3): 79-100.
- Chatterjee, Sayan and Birger Wernerfelt (1991). The link between resources and type of diversification: Theory and evidence. Strategic Management Journal, 12: 33-48.
- Cheung, Steven S. (1969). The Theory of Share Tenancy. Chicago, IL: Chicago University Press.
- Cheung, Steven S. (1973). The fable of the bees: An economic investigation. Journal of Law and Economics, 16: 11-34.
- Chi, Tailan (1994). Trading in strategic resources: Necessary conditions, transaction cost problems, and choice of exchange structure. Strategic Management Journal, 15: 271-290.

- Chi, Tailan (2000). Option to acquire or divest a joint venture. Strategic Management Journal, 21: 665-687.
- Clark, Kim B. and Takahiro Fujimoto (1991). Product Development Performance: Strategy, Organization, and Management in the World Auto Industry. Boston, MA: Harvard Business School Press.
- Coase, Ronald H. (1937). The nature of the firm. Economica, 4: 386-405.
- Coase, Ronald H. (1960). The problem of social cost. Journal of Law and Economics, 3: 1-44.
- Coase, Ronald H. (1988). The Firm, the Market and the Law. Chicago: University of Chicago Press.
- Coff, Russell W. (2002). Human capital, shared expertise, and the likelihood of impasse in corporate acquisitions. Journal of Management, 28 (1): 107-128.
- Cohen, Wesley M. and Daniel A. Levinthal (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35: 128-152.
- Coles, Jerilyn W. and William S. Hesterly (1998). The impact of firm-specific assets and the interaction of uncertainty: An examination of make or buy decisions on public and private hospitals. Journal of Economic Behavior and Organization, 36 (3): 383-409.
- Collis, David J. (1994). How valuable are organizational capabilities? Strategic Management Journal, 15: 143-152.
- Commons, John R. (1924). Legal Foundations of Capitalism. New York, NY: Macmillan.
- Commons, John R. (1934). Institutional Economics: Its Place in Political Economy. New York, NY: Macmillan.
- Conner, Kathleen R. (1991). A historical comparison of resource-based theory and five schools of thought within industrial organization economics: Do we have a new theory of the firm? Journal of Management, 17 (1): 121-154.
- Crocker, Keith J. and Scott E. Masten (1988). Mitigating contractual hazards: Unilateral options and contract length. Rand Journal of Economics 19: 327-343.
- Crocker, Keith J. and Kenneth J. Reynolds (1993). The efficiency of incomplete contracts: An empirical analysis of air force engine procurement. Rand Journal of Economics, 24 (1): 126-146.
- Cyert, Richard M., and James G. March (1963). A Behavioral Theory of the Firm. Englewood Cliffs, N.J.: Prentice-Hall.
- Davis, Lance E. and Douglass C. North (1971). Institutional Change and American Economic Growth. London: Cambridge University Press.
- Demsetz, Harold (1967). Toward a theory of property rights. American Economic Review, 57 (May): 347-359.

- Demsetz, Harold (1973). Industry structure, market rivalry, and public policy. Journal of Law and Economics, 16: 1-9.
- Demsetz, Harold (1988). Ownership, Control, and the Firm. Oxford: Basil Blackwell.
- Demsetz, Harold (1995). The Economics of the Business Firm. New York, NY: Cambridge University Press.
- Dierickx, Ingemar and Karel Cool (1989). Asset stock accumulation and sustainability of competitive advantage. Management Science, 35 (12): 1504-1511.
- Dixit, Avinash and Barry J. Nalebuff (1991). Thinking Strategically. New York, NY: W. W. Norton & Company.
- Dixit, Avinash and Robert S. Pindyck (1994). Investment Under Uncertainty. Princeton, NJ: Princeton University Press.
- Dosi, Giovanni, Richard R. Nelson and Sidney G. Winter (Eds.) (2000). The Nature and Dynamics of Organizational Capabilities. New York, NY: Oxford University Press.
- Duhaime, Irene M. and Charles R. Schwenk (1985). Conjectures on cognitive simplification in acquisition and divestment decision making. Academy of Management Review, 10: 287-295.
- Dyer, Jeffrey H. (1996). Specialized supplier networks as a source of competitive advantage: Evidence from the auto industry. Strategic Management Journal, 17: 271-291.
- Dyer, Jeffrey H. (1997). Effective interfirm collaboration: How firms minimize transaction costs and maximize transaction value. Strategic Management Journal, 18 (7): 535-556.
- Earl, Peter E. (2001). The Legacy of Herbert Simon in Economic Analysis. Cheltenham, UK: Edward Elgar.
- Eggertsson, Thrainn (1990). Economic Behavior and Institutions. Cambridge: Cambridge University Press.
- Eisenhardt, Kathleen M. (1985). Control: Organizational and economic approaches. Management Science, 31 (2): 134-149.
- Eisenhardt, Kathleen M. (1988). Agency- and institutional-theory explanations: The case of retail sales compensation. Academy of Management Journal, 31: 488-511.
- Eisenhardt, Kathleen M. (1989). Agency theory: An assessment and review. Academy of Management Review, 14 (1): 57-74.
- Evered, Roger and Meryl Reis Louis (1981). Alternative perspectives in the organizational sciences: "Inquiry from the inside" and "inquiry from the outside." Academy of Management Review, 6: 385-395.



- Fama, Eugene F. (1980). Agency problems and the theory of the firm. Journal of Political Economy, 88 (2): 288-307.
- Fama, Eugene F. and Michael C. Jensen (1983a). Separation of ownership and control. Journal of Law and Economics, 26: 301-325.
- Fama, Eugene F. and Michael C. Jensen (1983b). Agency problems and residual claims. Journal of Law and Economics, 26: 327-349.
- Farjoun, Moshe (1994). Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. Organization Science, 5 (2): 185-199.
- Farjoun, Moshe (1998). The independent and joint effects of the skill and physical bases of relatedness in diversification. Strategic Management Journal, 19: 611-630.
- Fiol, C. Marlene and Marjorie A. Lyles (1985). Organizational learning. Academy of Management Review, 10: 803-813.
- Folta, Timothy B. (1998). Governance and uncertainty: The trade-off between administrative control and commitment. Strategic Management Journal, 19: 1007-1028.
- Folta, Timothy B. and Kent D. Miller (2002). Real options in equity partnerships. Strategic Management Journal, 23: 77-88.
- Foss, Nicolai J. (1996). Knowledge-based approaches to the theory of the firm: Some critical comments. Organization Science, 7 (5): 470-476.
- Foss, Nicolai J. (Ed.) (1997). Resources, Firms and Strategies: A Reader in the Resource-Based Perspective. New York, NY: Oxford University Press.
- Friedman, Milton (1953). Essays in Positive Economics. Chicago, IL: University of Chicago Press.
- Fromm, Erich (1941). Escape from Freedom. New York, NY: Rinehart.
- Furubotn, Eirik, and Svetozar Pejovich (1972). Property rights and economic theory: A survey of recent literature. Journal of Economic Literature, 10: 1137-1162.
- Garud, Raghu and Arun Kumaraswamy (1995). Technological and organizational designs for realizing economies of substitution. Strategic Management Journal, 16 (Summer): 93-109.
- Garud, Raghu, Arun Kumaraswamy and Richard N. Langlois (Eds.), (2003). Managing in the Modular Age: Architectures, Networks and Organizations. New York, NY: Basil Blackwell.
- Ghemawat, Pankaj (1991). Commitment: The Dynamic of Strategy. New York, NY: Free Press.
- Ghemawat, Pankaj (1997). Games Businesses Play: Cases and Models. Cambridge, MA: MIT Press.

- Gimeno, Javier, Timothy B. Folta, Arnold C. Cooper and Carolyn Y. Woo (1997). Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. Administrative Science Quarterly, 42 (December): 750-783.
- Godfrey, Paul C. and Charles W. Hill (1995). The problem of unobservables in strategic management research. Strategic Management Journal, 16: 519-533.
- Goldberg, Victor P. and John R. Erickson (1987). Quantity and price adjustment in long-term contracts: A case study of petroleum coke. Journal of Law and Economics, 30: 369-398.
- Gort, Michael (1962). Diversification and Integration in American Industry. Princeton, NJ: Princeton University Press.
- Granovetter, Mark (1985). Economic action and social structure: The problem of embeddedness. American Journal of Sociology, 3: 481-510.
- Grant, Robert M. (1996). Toward a knowledge-based theory of the firm. Strategic Management Journal, 17 (December): 109-122.
- Grossman, Sanford J. and Oliver D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. Journal of Political Economy, 94: 691-719.
- Harrigan, Kathryn Rudie (1984). Formulating vertical integration strategies. Academy of Management Review, 9 (4): 638-652.
- Hart, Oliver (1989). An economist's perspective on the theory of the firm. Columbia Law Review, 89 (7): 1757-1774.
- Hart, Oliver (1995). Firms, Contracts, and Financial Structure. Oxford: Clarendon Press.
- Hart, Oliver and John Moore (1990). Property rights and the nature of the firm. Journal of Political Economy, 98: 1119-1158.
- Hayek, Friedrich A. (1945). The uses of knowledge in society. American Economic Review, 35: 519-530.
- Hayek, Friedrich A. (1948). Individualism and Economic Order. Chicago, IL: Henry Regnery Company, A Gateway Edition.
- Hayek, Friedrich A. (1978). New Studies in Philosophy, Politics, Economics and the History of Ideas. London, UK: Routledge & Kegan Paul.
- Heide, Jan B. and Anne S. Miner (1992). The shadow of the future: Effects of anticipated interaction and frequency of contract on buyer-supplier cooperation. Academy of Management Journal, 35 (2): 265-291.
- Heene, Aime and Ron Sanchez (Eds.) (1997). Competence-Based Strategic Management. New York, NY: John Wiley & Sons.

- Helfat, Constance E. (1997). Know-how and asset complementarity and dynamic capability accumulation: The case of R&D. Strategic Management Journal, 18 (5): 339-360.
- Helfat, Constance E. and Margaret A. Peteraf (2003). The dynamic resource-based view: Capability lifecycles. Strategic Management Journal 24: 997-1010.
- Henderson, Rebecca M. and Kim B. Clark (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative Science Quarterly, 35: 9-30.
- Henderson, Rebecca M. and Iain Cockburn (1994). Measuring competence? Exploring firm effects in pharmaceutical research. Strategic Management Journal, 15 (Winter): 63-84.
- Hennart, Jean-Francois (1982). A Theory of the Multinational Enterprise. Ann Arbor, MI: University of Michigan Press.
- Hennart, Jean-Francois (1988a). Upstream vertical integration in the aluminum and tin industries: A comparative study of the choice between market and intrafirm coordination. Journal of Economic Behavior and Organization, 9: 281-299.
- Hennart, Jean-Francois (1988b). A transaction costs theory of equity joint ventures. Strategic Management Journal, 9: 361-374.
- Hennart, Jean-Francois (1993). Explaining the swollen middle: Why most transactions are a mix of “market” and “hierarchy.” Organization Science, 4 (4): 529-547.
- Hirschman, Albert A. (1970). Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States. Cambridge, MA: Harvard University Press.
- Holmstrom, Bengt (1979). Moral hazard and observability. Bell Journal of Economics, 10: 74-91.
- Holmstrom, Bengt and John Roberts (1998). The boundaries of the firm revisited. Journal of Economic Perspectives 12 (4): 73-94.
- Hoopes, David G., Tammy L. Madsen and Gordon Walker (2003). Why is there a resource-based view? Toward a theory of competitive heterogeneity. Strategic Management Journal, 24: 889-902.
- Huff, Anne S. (1981). Multilectic methods of inquiry. Human Systems Management, 2: 83-94.
- Huff, Anne S. (2000). Changes in organizational knowledge production. Academy of Management Review, 25 (2): 288-293.
- Itami, Hiroyuki and Thomas E. Roehl (1987). Mobilizing Invisible Assets. Cambridge, MA: Harvard University Press.
- Jaffe, Jeffrey and James Michael Mahoney (1999). The performance of investment newsletters. Journal of Financial Economics 53 (2): 289-307.

- Jensen, Michael (1983). Organization theory and methodology. Accounting Review, 50 (April): 319-339.
- Jensen, Michael and William Meckling (1976). Theory of the firm: Managerial behavior, agency costs, and capital structure. Journal of Financial Economics, 3 (October): 305-360.
- John, George and Barton A. Weitz (1988). Forward integration into distribution: An empirical test of transaction cost analysis. Journal of Law, Economics and Organization, 4 (2): 337-355.
- Joskow, Paul L. (1985). Vertical integration and long-term contracts: The case of coal-burning electric generating plants. Journal of Law, Economics, and Organization, 1 (1): 33-80.
- Joskow, Paul L. (1987). Contract duration and relationship-specific investments: Empirical evidence from coal markets. American Economic Review, 77 (1): 1-25.
- Kahneman, Daniel, Paul Slovic and Amos Tversky (1982). Judgment Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.
- Kaplan, Abraham (1964). The Conduct of Inquiry: Methodology for Behavioral Science. San Francisco, CA: Chandler Publishing Company.
- Kim, Jongwook and Joseph T. Mahoney (2002). Resource-based and property rights perspectives on value creation: The case of oil field unitization. Managerial and Decision Economics, 23 (4): 225-245.
- Klein, Benjamin, Robert A. Crawford and Armen A. Alchian (1978). Vertical integration, appropriable rents, and the competitive contracting process. Journal of Law and Economics, 21 (October): 297-326.
- Klein, Benjamin and Keith B. Leffler (1981). The role of market forces in assuring contractual performance. Journal of Political Economy, 89 (4): 615-641.
- Knight, Frank H. (1921). Risk, Uncertainty and Profit. Boston, MA: Houghton Mifflin.
- Knott, Anne Marie (2003a). Persistent heterogeneity and sustainable innovation. Strategic Management Journal, 24: 687 – 705.
- Knott, Anne Marie (2003b). The organizational routines factor market paradox. Strategic Management Journal, 24: 929-943.
- Kogut, Bruce (1991). Joint ventures and the option to expand and acquire. Management Science, 37 (1): 19-33.
- Koza, Balaji and John E. Prescott (2002). Strategic alliances as social capital: A multidimensional view. Strategic Management Journal, 23: 795-816
- Kor, Yasemin Y. and Joseph T. Mahoney (2000). Penrose's resource-based approach: The process and product of research activity. Journal of Management Studies, 37 (1): 109-139.

- Kor, Yasemin Y. and Joseph T. Mahoney (2004). Edith Penrose's (1959) contributions to the resource-based view of strategic management. Journal of Management Studies, 41 (1): 183-191.
- Kosnik, Rita D. (1987). Greenmail: A study of board performance in corporate governance. Administrative Science Quarterly, 32: 163-185.
- Kuhn, Thomas S. (1970). The Structure of Scientific Revolutions. Chicago, IL: University of Chicago Press.
- Ladd, George W. (1987). Imagination in Research. Ames, Iowa: Iowa State Press.
- Lafontaine, Francine and Kathryn L. Shaw (1999). The dynamics of franchise contracting: Evidence from panel data. Journal of Political Economy, 107 (5): 1041-1080.
- Lajili, Kaouthar, Peter J. Barry, Steven T. Sonka and Joseph T. Mahoney (1997). Farmer's preferences for crop contracts. Journal of Agricultural and Resource Economics, 22 (2): 264-280.
- Langlois, Richard N. (1986). Economics as a Process: Essays in the New Institutional Economics. Cambridge, MA: Cambridge University Press.
- Langlois, Richard N. (2002). Modularity in technology and organization. Journal of Economic Organization, 49 (1): 19-37.
- Langlois, Richard N., Tony Fu-Lai Yu and Paul L. Robertson (Eds.) (2002). Alternative Theories of the Firm. Northampton, MA: Edward Elgar.
- Langlois, Richard N. and Paul L. Robertson (1995). Firms, Markets and Economic Change. London: Routledge.
- Leiblein, Michael J. and Douglas J. Miller (2003). An empirical examination of transaction- and firm-level influences on the vertical boundaries of the firm. Strategic Management Journal, 24: 839-859.
- Leiblein, Michael J., Jeffrey J. Reuer and Frederic Dalsace (2002). Do make or buy decisions matter? The influence of organizational governance on technological performance. Strategic Management Journal, 23: 817-833.
- Leonard-Barton, Dorothy (1992). Core capabilities and core rigidities: A paradox in managing new product development. Strategic Management Journal, 13 (Summer): 111-125.
- Leonard-Barton, Dorothy (1995). Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation. Boston, MA: Harvard Business School Press.
- Levinthal, Daniel (1988). A survey of agency models of organizations. Journal of Economic Behavior and Organization, 9: 153-185.
- Levy, David T. (1985). The transactions cost approach to vertical integration: An empirical examination. Review of Economics and Statistics, 67: 438-445.
- Libecap, Gary D. (1989). Contracting for Property Rights. Cambridge: Cambridge University Press.

- Libecap, Gary D. and Steven N. Wiggins (1985). The influence of private contractual failure on regulation: The case of oil field unitization. Journal of Political Economy, 93: 690-714.
- Lieberman, Marvin B. (1987). The learning curve, diffusion, and competitive strategy. Strategic Management Journal, 8: 441-452.
- Lieberman, Marvin B. (1991). Determinants of vertical integration: An empirical test. Journal of Industrial Economics, 39 (5): 451-466.
- Lieberman, Marvin B. and David B. Montgomery (1988). First-mover advantages. Strategic Management Journal, 9 (Summer): 41-58.
- Lieberman, Marvin B. and David B. Montgomery (1998). First-mover (dis)advantages: Retrospective and link with the resource-based view. Strategic Management Journal, 19: 1111-1125.
- Liebeskind, Julia Porter (1996). Knowledge, strategy, and the theory of the firm. Strategic Management Journal, 17 (December): 93-107.
- Lippman, Steven A. and Richard P. Rumelt (1982). Uncertain imitability: An analysis of interfirm differences in efficiency under competition. Bell Journal of Economics, 13: 418-438.
- Lippman, Steven A. and Richard P. Rumelt (2003). The payments perspective: Micro-foundations of resource analysis. Strategic Management Journal, 24: 903-927.
- Lyons, Bruce R. (1994). Contracts and specific investment: An empirical test of transaction costs theory. Journal of Economics and Management Strategy, 3 (2): 257-278.
- Machlup, Fritz (1967). Essays in Economic Semantics. New York, NY: W. W. Norton & Company.
- MacKinlay, A. Craig (1997). Event studies in economics and finance. Journal of Economic Literature, 35: 13-39.
- Macneil, Ian R. (1980). The New Social Contract: An Inquiry into Modern Contractual Relations. New Haven, CT: Yale University Press.
- Madhok, Anoop (2002). Reassessing the fundamentals and beyond: Ronald Coase, the transaction cost and resource-based theories of the firm and the institutional structure of production. Strategic Management Journal, 23: 535-550.
- Mahoney, James M. and Joseph T. Mahoney (1993). An empirical investigation of the effect of corporate charter antitakeover amendments on stockholder wealth. Strategic Management Journal, 14: 17-31.
- Mahoney, James M., Chamu Sundaramurthy, and Joseph T. Mahoney (1996). The differential impact on stockholder wealth of various antitakeover provisions. Managerial and Decision Economics, 17: 531-549.

- Mahoney, James M., Chamu Sundaramurthy, and Joseph T. Mahoney (1997). The effects of corporate antitakeover provisions on long-term investment: Empirical evidence. Managerial and Decision Economics, 18: 349-365.
- Mahoney, Joseph T. (1992a). Organizational economics within the conversation of strategic management. Advances in Strategic Management, 8: 103-155.
- Mahoney, Joseph T. (1992b). The adoption of the multidivisional form of organization: A contingency model. Journal of Management Studies, 29 (1): 49-72.
- Mahoney, Joseph T. (1992c). The choice of organizational form: Vertical financial ownership versus other methods of vertical integration. Strategic Management Journal, 13: 559-584.
- Mahoney, Joseph T. (1993). Strategic management and determinism: Sustaining the conversation. Journal of Management Studies, 30 (1): 173-191.
- Mahoney, Joseph T. (1995). The management of resources and the resource of management. Journal of Business Research, 33: 91-101.
- Mahoney, Joseph T. (2001). A resource-based theory of sustainable rents. Journal of Management 27: 651-660.
- Mahoney, Joseph T. (2002). The relevance of Chester I. Barnard's teachings to contemporary management education: Communicating the aesthetics of management. International Journal of Organization Theory and Behavior, 5 (1): 159-172.
- Mahoney, Joseph T. and David A. Crank (1995). Vertical coordination: The choice of organizational form. In Karen Coaldrake, Steven T. Sonka, Devanathan Sudharshan and Frederick W. Winter (Eds.), New Industries and Strategic Alliances in Agriculture: Concepts and Cases (pp. 89-126). Champaign, IL: Stipes Publishing.
- Mahoney, Joseph T., Anne S. Huff and James O. Huff (1994a). Toward a new social contract theory in organization science. Journal of Management Inquiry, 3 (2): 153-168.
- Mahoney, Joseph T., Anne S. Huff and James O. Huff (1994b). Management's search for balance. Journal of Management Inquiry, 3 (2): 173-174.
- Mahoney, Joseph T. and J. Rajendran Pandian (1992). The resource-based view within the conversation of strategic management. Strategic Management Journal, 13: 363-380.
- Mahoney, Joseph T. and Ron Sanchez (1997). Competence theory building: Reconnecting management research and management practice. In Aime Heene and Ron Sanchez (Eds.), Competence-based Strategic Management (pp. 45-64). Chichester, UK: John Wiley & Sons.
- Mahoney, Joseph T. and Ron Sanchez (2004). Building new management theory by integrating process and product of thought. Journal of Management Inquiry, forthcoming.
- Makadok, Richard (2001). Toward a synthesis of the resource-based and dynamic-capability views of rent creation. Strategic Management Journal, 22: 387-401.

- Makadok, Richard (2003). Doing the right thing and knowing the right thing to do: Why the whole is greater than the sum of the parts. Strategic Management Journal, 24: 1043-1055.
- Malmgren, Harold B. (1961). Information, expectations and the theory of the firm. Quarterly Journal of Economics, 75 (August): 399-421.
- March, James G. (1988). Decisions and Organizations. Oxford: Basil Blackwell.
- March, James G. (1999). The Pursuit of Organizational Intelligence. Oxford: Basil Blackwell.
- March, James G., and Herbert A. Simon (1958). Organizations. New York: John Wiley & Sons.
- Marris, Robin (1964). The Economic Theory of 'Managerial' Capitalism. New York, NY: Macmillan.
- Marschak, Jacob (1968). Economics of inquiring, communicating, deciding. American Economic Review, 58 (May): 1-18.
- Marshall, Alfred (1920). Principles of Economics. London: Macmillan and Company, Limited.
- Masten, Scott E. (1984). The organization of production: Evidence from the aerospace industry. Journal of Law and Economics, 27 (October): 403-418.
- Masten, Scott E. (1988). A legal basis for the firm. Journal of Law, Economics and Organization, 4: 181-198.
- Masten, Scott E. and Keith J. Crocker (1985). Efficient adaptation in long-term contracts: Take-or-pay provision for natural gas. American Economic Review, 75: 1083-1093.
- Masten, Scott E., James W. Meehan, and Edward A. Snyder (1989). Vertical integration in the U.S. auto industry: A note on the influence of transaction specific assets. Journal of Economic Behavior and Organization, 12: 265-273.
- Masten, Scott E., James W. Meehan, and Edward A. Snyder (1991). The costs of organization. Journal of Law, Economics and Organization, 7 (1): 1-25.
- McCloskey, Donald (1993). The rhetoric of economics. Journal of Economic Literature, 21 (June): 481-517.
- McCloskey, Deidre (1998). The Rhetoric of Economics. Madison, WI: University of Wisconsin Press.
- McCloskey, D. and Stephen T. Ziliak (1996). The standard error of regressions. Journal of Economic Literature, 34 (March): 97-114.
- McGrath, Rita Gunther (1997). A real options logic for initiating technology positioning investments. Academy of Management Review, 22 (4): 974-996.



- McGrath, Rita Gunther (1999). Falling forward: Real options reasoning and entrepreneurial failure. Academy of Management Review, 24 (1): 13-30.
- McMillan, John (1992). Games, Strategies, and Managers. New York, NY: Oxford University Press.
- Michael, Steven C. (2000). Investments to create bargaining power: The case of franchising. Strategic Management Journal, 21: 497-514.
- Milgrom, Paul and John Roberts (1992). Economics, Organization and Management. Englewood Cliffs, NJ: Prentice Hall.
- Miller, Danny and Jamal Shamsie (1996). The resource-based view of the firm in two environments: The Hollywood film studios from 1936 to 1965. Academy of Management Journal, 39 (3); 519-534.
- Miller, Gary J. (1992). Managerial Dilemmas: The Political Economy of Hierarchy. New York, NY: Cambridge University Press.
- Miller, Kent D. (2002). Knowledge inventories and managerial myopia. Strategic Management Journal, 23: 689-706.
- Miller, Kent D. and Timothy B. Folta (2002). Option value and entry timing. Strategic Management Journal, 23: 655-665.
- Monteverde, Kirk (1995). Technical dialog as an incentive for vertical integration in the semiconductor industry. Management Science, 41 (10): 1624-1638.
- Monteverde, Kirk and David Teece (1982). Supplier switching costs and vertical integration in the automobile industry. Bell Journal of Economics, 13 (Spring): 206-213.
- Montgomery, Cynthia A. and S. Hariharan (1991). Diversified expansion by large established firms. Journal of Economic Behavior and Organization, 15: 71-89.
- Montgomery, Cynthia A. and Birger Wernerfelt (1988). Diversification, Ricardian rents, and Tobin's q. Rand Journal of Economics, 19 (4):623-632.
- Montgomery, Cynthia A., Birger Wernerfelt and Srinivasan Balakrishnan (1989). Strategy content and the research process. Strategic Management Journal, 10: 189-197.
- Mosakowski, Elaine (1991). A resource-based perspective on the dynamic strategy-performance relationship: An empirical examination of the focus and differentiation strategies in entrepreneurial firms. Journal of Management, 19 (4): 819-839.
- Mowery, David C., Joanne E. Oxley and Brian S. Silverman (1996). Strategic alliances and interfirm knowledge transfer. Strategic Management Journal, 17 (Winter): 77-91.
- Nelson, Richard R. (1996). The Sources of Economic Growth. Cambridge, MA: Harvard University Press.

- Nelson, Richard R., and Sidney G. Winter (1982). An Evolutionary Theory of Economic Change. Cambridge, Mass.: Harvard University Press.
- Nelson, Richard R., and Sidney G. Winter (2002). Evolutionary theorizing in economics. Journal of Economic Perspectives, 16 (2): 23-46.
- Nichols, Nancy A. (1994). Scientific management at Merck: An interview with CFO Judy Lewent. Harvard Business Review (January-February): 89-99.
- Nickerson, Jack A., Barton H. Hamilton and Tetsuo Wada (2001). Market position, resource profile, and governance: Linking Porter and Williamson in the context of international courier and small package services in Japan. Strategic Management Journal, 22: 251-273.
- Nonaka, Ikujiro and Hirotaka Takeuchi (1995). The Knowledge-Creating Company. Oxford: Oxford University Press.
- North, Douglass (1981). Structure and Change in Economic History. New York, NY: W. W. Norton.
- North, Douglass C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.
- North, Douglass C. and Robert P. Thomas (1973). The Rise of the Western World: A New Economic History. Cambridge: Cambridge University Press.
- Norton, Seth W. (1988). Franchising, brand name capital, and the entrepreneurial capacity problem. Strategic Management Journal, 9: 105-114.
- Novak, Sharon and Steven D. Eppinger (2001). Sourcing by design: Product complexity and the supply chain. Management Science, 47 (1): 189-204.
- O'Driscoll, Gerald P. and Mario J. Rizzo (1985). The Economics of Time and Ignorance. New York, NY: Basil Blackwell.
- Oliver, Christine (1997). Sustainable competitive advantage: Combining institutional and resource-based views. Strategic Management Journal, 18: 697-713.
- Ouchi, William G. (1979). A conceptual framework for the design of organizational control mechanisms. Management Science, 25 (9): 833-848.
- Oviatt, Benjamin M. (1988). Agency and transaction cost perspectives on the manager-shareholder relationship: Incentives for congruent interests. Academy of Management Review, 13: 214-225.
- Palay, Thomas M. (1984). Comparative institutional economics: The governance of rail freight contracting. Journal of Legal Studies, 13: 265-287.
- Palay, Thomas M. (1985). Avoiding regulatory constraints: The use of informal contracts. Journal of Law, Economics, and Organization, 1 (Spring): 155-175.

- Palmer, Donald, Roger Friedland, P. Devereaux Jennings and Melanie E. Powers (1987). The economics and politics of structure: The multidivisional form and the large U.S. corporation. Administrative Science Quarterly, 32: 25-48.
- Peng, Mike W. (2001). The resource-based view and international business. Journal of Management, 27: 803-829.
- Penrose, Edith T. (1955). Limits to the growth of the firm. American Economic Review, 45 (2): 531-543.
- Penrose, Edith T. (1959). The Theory of the Growth of the Firm. New York: John Wiley & Sons.
- Penrose, Edith T. (1960). The growth of the firm: A case study: The Hercules Powder Company. Business History Review, 34: 1-23.
- Peteraf, Margaret (1993). The cornerstones of competitive advantage: A resource-based view. Strategic Management Journal, 14: 179-191.
- Pettus, Michael (2001). The resource-based view as a developmental growth process: Evidence from the deregulated trucking industry. Academy of Management Journal, 44 (4): 878-896.
- Phillips, Almarin and Joseph Mahoney (1985). Unreasonable rules and rules of reason: Economic aspects of vertical price-fixing. Antitrust Bulletin, 30 (1): 99-115.
- Pirrong, Stephen C. (1993). Contracting practices in bulk shipping markets: A transaction costs explanation. Journal of Law and Economics, 36: 937-976.
- Pisano, Gary P. (1989). Using equity participation to support exchange: Evidence from the biotechnology industry. Journal of Law, Economics and Organization, 5 (1): 109-126.
- Pitelis, Christos (Ed.) (2002). The Growth of the Firm: The Legacy of Edith Penrose. New York, NY: Oxford University Press.
- Polanyi, Michael (1962). Personal Knowledge. Chicago, IL: University of Chicago Press.
- Popper, Karl R. (1970). Normal science and its dangers. In L. Lakatos and A. Musgrave (Eds.), Criticism and the Growth of Knowledge. Cambridge: Cambridge University Press.
- Poppo, Laura and Todd Zenger (1998). Testing alternative theories of the firm: Transaction cost, knowledge-based, and measurement explanations for make-or-buy-decisions in information services. Strategic Management Journal, 19: 853-877.
- Poppo, Laura and Todd Zenger (2002). Do formal contracts and relational governance function as substitutes or complements? Strategic Management Journal, 23: 707-725.
- Porter, Michael E. (1980). Competitive Strategy: Techniques for Analyzing Industries and Competitors. New York, NY: Free Press.

- Porter, Michael E. (1985). Competitive Advantage: Creating and Sustaining Superior Performance. New York, NY: Free Press.
- Porter, Michael E. (1991). Towards a dynamic theory of strategy. Strategic Management Journal, 12: 95-117.
- Powell, Thomas C. and Anne Dent-Micallef (1997). Information technology as competitive advantage: The role of human, business, and technology resources. Strategic Management Journal, 18 (5): 375-405.
- Pratt, John W., and Richard J. Zeckhauser (1985). "Principals and Agents: An Overview" (pp. 1-35). In Principals and Agents: The Structure of Business, edited by John W. Pratt and Richard J. Zeckhauser. Boston, MA: Harvard Business School Press.
- Rasmusen, Eric (1989). Games and Information. Cambridge, MA: Basil Blackwell.
- Rediker, Kenneth J. and Anju Seth (1995). Boards of directors and substitution effects of alternative governance mechanisms. Strategic Management Journal, 16 (2): 85-99.
- Redman, Deborah A. (1993). Economics and the Philosophy of Science. New York, NY: Oxford University Press.
- Reed, Richard and Robert J. DeFillippi (1990). Causal ambiguity, barriers to imitation and sustainable competitive advantage. Academy of Management Review, 15 (1): 88-102.
- Ricardo, David (1817). Principles of Political Economy and Taxation. London: J. Murray.
- Richardson, George B. (1960). Information and Investment. Oxford, UK: Oxford University Press.
- Richardson, George B. (1972). The organization of industry. Economic Journal, 82 (September): 883-896.
- Richardson, James (1993). Parallel sourcing and supplier performance in the Japanese automobile industry. Strategic Management Journal, 14: 339-350.
- Riordan, Michael and Oliver E. Williamson (1985). Asset specificity and economic organization. International Journal of Industrial Organization, 3: 365-378.
- Robertson, Thomas S. and Hubert Gatignon (1998). Technology development mode: A transaction cost conceptualization. Strategic Management Journal, 19: 515-531.
- Rubin, Paul H. (1973). The expansion of firms. Journal of Political Economy, 81: 936-949.
- Rubin, Paul H. (1978). The theory of the firm and the structure of the franchise contract. Journal of Law and Economics, 21: 223-233.
- Rugman, Alan M. and Alain Verbeke (2002). Edith Penrose's contribution to the resource-based view of strategic management. Strategic Management Journal, 23: 769-780.

- Rumelt, Richard P. (1974). Strategy, Structure, and Economic Performance. Boston, MA: Harvard Business School Press.
- Rumelt, Richard P. (1984). Toward a strategic theory of the firm. In R. Lamb (Ed.), Competitive Strategic Management (pp. 556-570). Englewood Cliffs, NJ: Prentice-Hall.
- Rumelt, Richard P., Dan E. Schendel, and David J. Teece (Eds.) (1994). Fundamental Issues in Strategy. Boston, MA: Harvard University Press.
- Saloner, Garth 1991, "Modeling, game theory, and strategic management," *Strategic Management Journal*, Vol. 12 (Winter), pp. 119-136.
- Sanchez, Ron (1993). Strategic flexibility, firm organization, and managerial work in dynamic markets: A strategic options perspective. Advances in Strategic Management, 9: 251-291.
- Sanchez, Ron (2003). Integrating transaction costs theory and real options theory. Managerial and Decision Economics, 24: 267-282.
- Sanchez, Ron and Joseph T. Mahoney (1996). Modularity, flexibility, and knowledge management in product and organization design. Strategic Management Journal (December): 63-76.
- Sanchez, Ron and Joseph T. Mahoney (2001). Modularity and dynamic capabilities. In Henk Volberda and Tom Elfring (Eds.) Rethinking Strategy (pp. 158-171). Thousand Oaks, CA: Sage Publications.
- Schelling, Thomas C. (1960). The Strategy of Conflict. Cambridge, MA: Harvard University Press.
- Scherer, Frederick M. and David Ross (1990). Industrial Market Structure and Economic Performance. Boston, MA: Houghton Mifflin Company.
- Schilling, Melissa A. (2000). Toward a general modular systems theory and its application to interfirm product modularity. Academy of Management Review, 25 (2): 312-334.
- Schoemaker, Paul, J. H. (1990). Strategy, complexity and economic rent. Management Science, 36: 1178-1192.
- Schumpeter, Joseph A. (1934). The Theory of Economic Development. Cambridge, MA: Harvard University Press.
- Schumpeter, Joseph A. (1950). Capitalism, Socialism, and Democracy. New York, NY: Harper & Brothers Publishers.
- Schwartz, Eduardo S. and Lenos Trigeorgis (Eds.) (2001). Real Options and Investment Under Uncertainty: Classical Readings and Recent Contributions. Cambridge, MA: MIT Press.
- Scott, W. Richard (1987). Organizations: Rational, Natural, and Open Systems. Englewood Cliffs, NJ: Prentice-Hall.
- Scott, W. Richard (1995). Institutions and Organizations. Thousand Oaks, CA: Sage Publications.

- Selznick, Philip (1957). Leadership in Administration. Berkeley, CA: University of California Press.
- Seth, Anju (1990). Sources of value creation in acquisitions: An empirical investigation. Strategic Management Journal, 11: 431-446.
- Seth, Anju and Howard Thomas (1994). Theories of the firm: Implications for strategy research. Journal of Management Studies, 31 (2): 165-191.
- Seth, Anju and George Zinkhan (1991). Strategy and the research process: A comment. Strategic Management Journal, 12: 75-82.
- Shane, Scott (1996). Hybrid organizational arrangements and their implications for firm growth and survival: A study of new franchisors. Academy of Management Journal, 39 (1): 216-234.
- Shapiro, Carl and Hal R. Varian (1999). Information Rules: A Strategic Guide to the Network Economy. Boston, MA: Harvard Business School Press.
- Silverman, Brian S. (1999). Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. Management Science, 45: 1109-1124.
- Simon, Herbert A. (1947). Administrative Behavior. New York, NY: Macmillan.
- Simon, Herbert A. (1957). Models of Man: Social and Rational. New York, NY: John Wiley & Sons.
- Simon, Herbert A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106: 467-482.
- Simon, Herbert A. (1982). Models of Bounded Rationality: Behavioral Economics and Business Organization. Cambridge, Mass.: MIT Press.
- Simon, Herbert A. (1996). The Sciences of the Artificial. Cambridge, MA: MIT Press.
- Slater, Martin (1980). The managerial limitations to the growth of firm. Economic Journal 90: 520-528.
- Smith, Adam (1776). The Wealth of Nations. New York, NY: Modern Library Edition of 1937, edited by Edwin Cannan.
- Sorenson, Olav and Jesper B. Sorensen (2001). Finding the right mix: Franchising, organizational learning, and chain performance. Strategic Management Journal, 22: 713-724.
- Spender, J.-C. (1996). Making knowledge the basis of a dynamic theory of the firm. Strategic Management Journal, 17 (December): 45-62.
- Stigler, George J. (1968). The Organization of Industry. Homewood, IL: Irwin.
- Stuckey, John (1983). Vertical Integration and Joint Ventures in the Aluminum Industry. Cambridge, MA: Harvard University Press.

- Subramani, Mani R. and N. Venkatraman (2003). Safeguarding investments in asymmetric interorganizational relationships: Theory and evidence. Academy of Management Journal, 46 (1): 46-62.
- Sundaramurthy, Chamu and Marianne Lewis (2003). Control and collaboration: Paradoxes of governance. Academy of Management Review, 28 (3): 397-415.
- Sundaramurthy, Chamu, James M. Mahoney and Joseph T. Mahoney (1997). The effects of corporate antitakeover provisions on long-term investment: Empirical evidence. Strategic Management Journal, 18 (3): 231-245.
- Sutton, John (1992). Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration. Cambridge, MA: MIT Press.
- Szulanski, Gabriel (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. Strategic Management Journal, 17 (Winter): 27-43.
- Szulanski, Gabriel (2003). Sticky Knowledge: Barriers to Knowing in the Firm. Thousand Oaks, CA: Sage Publications.
- Tan, Danchi. (2003). The limits to the growth of multinational firms in a foreign market. Managerial and Decision Economics, 24: 569-582.
- Tan, Danchi and Joseph T. Mahoney (2003). Explaining the utilization of managerial expatriates from the perspectives of resource-based, agency, and transaction costs theories. Advances in International Management, 15: 179-205.
- Tan, Danchi and Joseph T. Mahoney (2004). Examining the Penrose effect in an international business context: The dynamics of Japanese growth in U.S. industries. Managerial and Decision Economics, forthcoming.
- Teece, David J. (1980). Economies of scope and the scope of the enterprise. Journal of Economic Behavior and Organization, 1: 223-245.
- Teece, David J. (1982). Towards an economic theory of the multiproduct firm. Journal of Economic Behavior and Organization, 3: 39-63.
- Teece, David J. (2000). Managing Intellectual Capital: Organizational, Strategic and Policy Dimensions. New York, NY: Oxford University Press.
- Teece, David J., Gary Pisano and Amy Shuen (1997). Dynamic capabilities and strategic management. Strategic Management Journal, 18 (7): 509-533.
- Telser, Lester. G. (1980). A theory of self-enforcing agreements. Journal of Business, 53: 27-44.
- Thomke, Stefan and Walter Kuemmerle (2002). Asset accumulation, interdependence and technological change: Evidence from pharmaceutical drug discovery. Strategic Management Journal, 23: 619-635.
- Thompson, James D. (1967). Organizations in Action. New York, NY: McGraw-Hill.

- Tirole, Jean (1988). The Theory of Industrial Organization. Cambridge, MA: MIT Press.
- Trigeorgis, Lenos (1997). Real Options: Managerial Flexibility and Strategy in Resource Allocation. Cambridge, MA: MIT Press.
- Tripsas, Mary (1997). Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. Strategic Management Journal, 18 (Summer): 119-142.
- Volberda, Henk W. and Tom Elfring (Eds.) (2001). Rethinking Strategy. Thousand Oaks, CA: Sage Publications.
- Walker, Gordon and Laura Poppo (1991). Profit centers, single-source suppliers, and transaction costs. Administrative Science Quarterly, 36: 66-87.
- Walker, Gordon and David Weber (1984). A transaction cost approach to make-or-buy decisions. Administrative Science Quarterly, 29 (September): 373-391.
- Walker, Gordon and David Weber (1987). Supplier competition, uncertainty and make-or-buy decisions. Academy of Management Journal, 30: 589-596.
- Walsh, James P. and James K. Seward (1990). On the efficiency of internal and external corporate control mechanisms. Academy of Management Review, 15 (3): 421-458.
- Wernerfelt, Birger (1984). A resource-based view of the firm. Strategic Management Journal, 5: 171-180.
- Wernerfelt, Birger (1995). The resource-based view of the firm: Ten years after. Strategic Management Journal, 171-174.
- Whetten, David A. (1989). What constitutes a theoretical contribution? Academy of Management Review, 14 (4): 490-495.
- Williamson, Oliver E. (1971). The vertical integration of production: Market failure considerations. American Economic Review, 61 (May): 112-123.
- Williamson, Oliver E. (1975). Markets and Hierarchies: Analysis and Antitrust Implications. New York, NY: Free Press.
- Williamson, Oliver E. (1979). Transaction cost economics: The governance of contractual relations. Journal of Law and Economics, 22: 233-261.
- Williamson, Oliver E. (1985). The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. New York, NY: The Free Press.
- Williamson, Oliver E. (1991). Strategizing, economizing, and economic organization. Strategic Management Journal, 12 (Winter): 75-94.



- Williamson, Oliver E. (Ed.) (1995). Organization Theory: From Chester Barnard to the Present and Beyond. Oxford: Oxford University Press.
- Williamson, Oliver E. (1996). The Mechanisms of Governance. New York, NY: Oxford University Press.
- Williamson, Oliver E. (1999). Strategy research: Governance and competence perspectives. Strategic Management Journal, 20: 1087-1108.
- Williamson, Oliver E. (2000). The new institutional economics: Taking stock and looking ahead. Journal of Economic Literature, 38 (September): 596-613.
- Williamson, Oliver E. and Scott E. Masten (Eds.) (1999). The Economics of Transaction Costs. Northampton, MA: Elgar.
- Williamson, Oliver E. and Sidney G. Winter (Eds.) (1991). The Nature of the Firm: Origin, Evolution, Development. New York, NY: Oxford University Press.
- Wolfson, M. A. (1985). Empirical evidence of incentive problems and their mitigation in oil and gas shelter programs (pp. 101 – 125). In J.W. Pratt and R. J. Zeckhauser (Eds.), Principals and Agents, Boston, MA: Harvard Business School Press.
- Worren, Nicolay, Karl Moore and Pablo Cardona (2002). Modularity, strategic flexibility, and firm performance: A study of the home appliance industry. Strategic Management Journal, 23: 1123-1140.
- Yao, Dennis A. (1988). Beyond the reach of the invisible hand: Impediments to economic activity, market failures and profitability. Strategic Management Journal, 9 (Summer): 59-70.
- Yeoh, Poh-Lin and Kendall Roth (1999). An empirical analysis of sustained advantage in the U.S. pharmaceutical industry: Impact of firm resources and capabilities. Strategic Management Journal, 20: 637-653.
- Zander, Udo and Bruce Kogut (1995). Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. Organization Science, 6 (1): 76-92.
- Zenger, Todd R. (1994). Explaining organizational diseconomies of scale in R&D: Agency problems and the allocation of engineering talent, ideas, and effort by firm size. Management Science, 40 (6): 708-729.
- Zingales, Luigi (2000). In search of new foundations. Journal of Finance, 55: 1623-1653.

## Biographical Sketch

**Joseph T. Mahoney** received his BA in Economics at the University of Pennsylvania in 1980, his Masters (in 1984), and his Ph.D. (in 1989) in Business Economics at the Wharton School of Business of the University of Pennsylvania. Joe joined the faculty within the Department of Business Administration of the College of Business at University of Illinois at Urbana-Champaign in January 1988. Joe was promoted to Associate Professor in 1995 and he was promoted to Full Professor in 2003. Currently, Joe is on the editorial board of *Academy of Management Review*, *Journal of Management Studies*, and *Strategic Management Journal*. For the 2000-2002 period, he was the book review editor of *Academy of Management Review*. He has published in *Academy of Management Review*, *Journal of Business Research*, *Journal of Management*, *Journal of Management Inquiry*, *Journal of Management Studies*, *Managerial and Decision Economics*, and *Strategic Management Journal*, among others.