Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction. In consequence they structure incentives in human exchange, whether political, social, or economic. Institutional change shapes the way societies evolve through time and hence is the key to understanding historical change.

That institutions affect the performance of economies is hardly controversial. That the differential performance of economies over time is fundamentally influenced by the way institutions evolve is also not controversial. Yet neither current economic theory nor cliometric history shows many signs of appreciating the role of institutions in economic performance because there as yet has been no analytical framework to integrate institutional analysis into economics and economic history. The objective of this book is to provide such an underlying framework. The implications of the analysis suggest a reexamination of much social science theorizing in general and economics in particular, and provide a new understanding of historical change.

In this study I examine the nature of institutions and the consequences of institutions for economic (or societal) performance (Part I). I then outline a theory of institutional change not only to provide a framework for economic (and other) history, but also to explain how the past influences the present and future, the way incremental institutional change affects the choice set at a moment of time, and the nature of path dependence (Part II). The primary objective of the study is to achieve an understanding of the differential performance of economies through time (Part III).
**Institutions**

friends on the street, drive an automobile, buy oranges, borrow money, form a business, bury our dead, or whatever, we know (or can learn easily) how to perform these tasks. We would readily observe that institutions differ if we were to try to make the same transactions in a different country – Bangladesh for example. In the jargon of the economist, institutions define and limit the set of choices of individuals.

Institutions include any form of constraint that human beings devise to shape human interaction. Are institutions formal or informal? They can be either, and I am interested both in formal constraints – such as rules that human beings devise – and in informal constraints – such as conventions and codes of behavior. Institutions may be created, as was the United States Constitution; or they may simply evolve over time, as does the common law. I am interested in both created and evolving institutions, although for purposes of analysis we may want to examine them separately. Many other attributes of institutions also will be explored.

Institutional constraints include both what individuals are prohibited from doing and, sometimes, under what conditions some individuals are permitted to undertake certain activities. As defined here, they therefore are the framework within which human interaction takes place. They are perfectly analogous to the rules of the game in a competitive team sport. That is, they consist of formal written rules as well as typically unwritten codes of conduct that underlie and supplement formal rules, such as not deliberately injuring a key player on the opposing team. And as this analogy would imply, the rules and informal codes are sometimes violated and punishment is enacted. Therefore, an essential part of the functioning of institutions is the costliness of ascertaining violations and the severity of punishment.

Continuing the sports analogy, taken together, the formal and informal rules and the type and effectiveness of enforcement shape the whole character of the game. Some teams are successful as a consequence of (and have therefore the reputation for) constantly violating rules and thereby intimidating the opposing team. Whether that strategy pays off obviously depends on the effectiveness of monitoring and the severity of punishment. Sometimes codes of conduct – good sportsmanship – constrain players, even though they could get away with successful violations.

A crucial distinction in this study is made between institutions and organizations. Like institutions, organizations provide a structure to human interaction. Indeed when we examine the costs that arise as a consequence of the institutional framework we see they are a result not only of that framework, but also of the organizations that have developed in consequence of that framework. Conceptually, what must be clearly differentiated are the rules from the players. The purpose of the rules is to define the way the game is played. But the objective of the team within
An introduction

that set of rules is to win the game — by a combination of skills, strategy, and coordination; by fair means and sometimes by foul means. Modeling the strategies and the skills of the team as it develops is a separate process from modeling the creation, evolution, and consequences of the rules.

Organizations include political bodies (political parties, the Senate, a city council, a regulatory agency), economic bodies (firms, trade unions, family farms, cooperatives), social bodies (churches, clubs, athletic associations), and educational bodies (schools, universities, vocational training centers). They are groups of individuals bound by some common purpose to achieve objectives. Modeling organizations is analyzing governance structures, skills, and how learning by doing will determine the organization’s success over time. Both what organizations come into existence and how they evolve are fundamentally influenced by the institutional framework. In turn they influence how the institutional framework evolves. But as noted above, the emphasis in this study is on the institutions that are the underlying rules of the game and the focus on organizations (and their entrepreneurs) is primarily on their role as agents of institutional change; therefore the emphasis is on the interaction between institutions and organizations. Organizations are created with purposive intent in consequence of the opportunity set resulting from the existing set of constraints (institutional ones as well as the traditional ones of economic theory) and in the course of attempts to accomplish their objectives are a major agent of institutional change.

Separating the analysis of the underlying rules from the strategy of the players is a necessary prerequisite to building a theory of institutions. Defining institutions as the constraints that human beings impose on themselves makes the definition complementary to the choice theoretic approach of neoclassical economic theory. Building a theory of institutions on the foundation of individual choices is a step toward reconciling differences between economics and the other social sciences. The choice theoretic approach is essential because a logically consistent, potentially testable set of hypotheses must be built on a theory of human behavior. The strength of microeconomic theory is that it is constructed on the basis of assumptions about individual human behavior (even though I shall argue for a change in those assumptions in Chapter 3). Institutions are a creation of human beings. They evolve and are altered by human beings; hence our theory must begin with the individual. At the same time, the constraints that institutions impose on individual choices are pervasive. Integrating individual choices with the constraints institutions impose on choice sets is a major step toward unifying social science research.

Institutions affect the performance of the economy by their effect on the costs of exchange and production. Together with the technology em-
Institutions

ployed, they determine the transaction and transformation (production) costs that make up total costs. The initial objective of this study (Part I) is to explain the existence and nature of institutions to specify the way they enter into the cost functions in an economy.

II

The major role of institutions in a society is to reduce uncertainty by establishing a stable (but not necessarily efficient) structure to human interaction. But the stability of institutions in no way gainsays the fact that they are changing. From conventions, codes of conduct, and norms of behavior to statute law, and common law, and contracts between individuals, institutions are evolving and, therefore, are continually altering the choices available to us. The changes at the margin may be so slow and glacial in character that we have to stand back as historians to perceive them, although we live in a world where the rapidity of institutional change is very apparent.

Institutional change is a complicated process because the changes at the margin can be a consequence of changes in rules, in informal constraints, and in kinds and effectiveness of enforcement. Moreover, institutions typically change incrementally rather than in discontinuous fashion. How and why they change incrementally and why even discontinuous changes (such as revolution and conquest) are never completely discontinuous are a result of the imbeddedness of informal constraints in societies. 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 provide us with a key to explaining the path of historical change.

The central puzzle of human history is to account for the widely divergent paths of historical change. How have societies diverged? What accounts for their widely disparate performance characteristics? After all, we all descended from primitive hunting and gathering bands. This divergence is even more perplexing in terms of standard neoclassical and international trade theory, which implies that over time economies, as they traded goods, services, and productive factors, would gradually converge. 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 religious, ethnic, cultural, political, and economic societies, and the gap between rich and poor nations, between developed and undeveloped nations, is as wide today as it ever was and perhaps a great deal wider than ever before. What
An introduction

explains the divergence? And perhaps equally important, what conditions either lead to further divergences or produce convergence?

There is more to this puzzle. What accounts for societies experiencing long-run stagnation or an absolute decline in economic well-being? The evolutionary hypothesis advanced by Alchian in 1950 would suggest that ubiquitous competition would weed out inferior institutions and reward by survival those that better solve human problems.

Let me briefly retrace my steps in dealing with this central issue. In North and Thomas (1973) we made institutions the determinant of economic performance and relative price changes the source of institutional change. But we had an essentially efficient explanation; changes in relative prices create incentives to construct more efficient institutions. The persistence of inefficient institutions, illustrated by the case of Spain, was a result of fiscal needs of rulers that led to shortened time horizons and therefore a disparity between private incentives and social welfare. Such an anomaly did not fit into the theoretical framework.

In Structure and Change in Economic History (North, 1981) I abandoned the efficiency view of institutions. Rulers devised property rights in their own 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. In that study I raised the question posed by Alchian’s evolutionary argument, but had no answer. It was possible to explain the existence of inefficient institutions, but why wouldn’t competitive pressures lead to their elimination? Wouldn’t the political entrepreneurs in stagnant economies quickly emulate the policies of more successful ones? How can we explain the radically differential performance of economies over long periods of time?

This study answers these questions. The answer hinges on the difference between institutions and organizations and the interaction between them that shapes the direction of institutional change. Institutions, together with the standard constraints of economic theory, determine the opportunities in a society. Organizations are created to take advantage of those opportunities, and, as the organizations evolve, they alter the institutions. The resultant path of institutional change is shaped by (1) the lock-in that comes from the symbiotic relationship between institutions and the organizations that have evolved as a consequence of the incentive structure provided by those institutions and (2) the feedback process by which human beings perceive and react to changes in the opportunity set.

The increasing returns characteristics of an institutional matrix that produces lock-in come from the dependence of the resultant organizations on that institutional framework and the consequent network exter-
Institutions

Institutions that arise. Both the formal and the informal institutional constraints result in particular exchange organizations that have come into existence because of the incentives embodied in the framework and therefore depend on it for the profitability of the activities that they undertake.

Incremental change comes from the perceptions of the entrepreneurs in political and economic organizations that they could do better by altering the existing institutional framework at some margin. But the perceptions crucially depend on both the information that the entrepreneurs receive and the way they process that information. If political and economic markets were efficient (i.e., there were zero transaction costs) then the choices made would always be efficient. That is the actors would always possess true models or if they initially possessed incorrect models the information feedback would correct them. But that version of the rational actor model has simply led us astray. The actors frequently must act on incomplete information and process the information that they do receive through mental constructs that 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 players as they attempt to understand the complexities of the problems they confront can lead to the persistence of such property rights.

We can expand on this characterization of institutional change by contrasting a successful path with one of persistent failure. The first is a familiar story in U.S. economic history—the growth of the economy in the nineteenth century. The basic institutional framework that had evolved by the beginning of that century (the Constitution and the Northwest Ordinance, as well as norms of behavior rewarding hard work) broadly induced the development of economic and political organizations (Congress, local political bodies, family farms, merchant houses, and shipping firms), whose maximizing activities resulted in increased productivity and economic growth both directly and indirectly by an induced demand for educational investment. The educational investment resulted not only in the free public educational system, but in agricultural experiment stations to improve agricultural productivity; the Morrill Act created the land grant public universities.

As economic organizations evolved to take advantage of these opportunities, they not only became more efficient (see Chandler, 1977), but also gradually altered the institutional framework. Not only was the political and judicial framework altered (the Fourteenth Amendment, *Munn v. Illinois*) and the structure of property rights modified (the Sherman Act) by the end of the nineteenth century, but so too were many norms of behavior and other informal constraints (reflected in changing attitudes—and norms of behavior—toward slavery, the role of women, and temperance, for example). Both the political and the economic trans-
An introduction

action costs and the subjective perceptions of the actors resulted in choices that were certainly not always optimal or unidirectional toward increased productivity or improved economic welfare (however defined). The profitable opportunities were sometimes from tariff creation, the exploitation of slaves, or the formation of a trust. Sometimes, indeed frequently, policies had unintended consequences. In consequence institutions were — and are — always a mixed bag of those that induce productivity increase and those that reduce productivity. Institutional change, likewise, almost always creates opportunities for both types of activity. But on balance nineteenth-century U.S. economic history is a story of economic growth because the underlying institutional framework persistently reinforced incentives for organizations to engage in productive activity however admixed with some adverse consequences.

Now if I describe an institutional framework with a reverse set of incentives to those described in the above paragraph, I will approximate the conditions in many Third World countries today as well as those that have characterized much of the world's economic history. The opportunities for political and economic entrepreneurs are still a mixed bag, but they overwhelmingly favor activities that promote redistributive rather than productive activity, that create monopolies rather than competitive conditions, and that restrict opportunities rather than expand them. They seldom induce investment in education that increases productivity. The organizations that develop in this institutional framework will become more efficient — but more efficient at making the society even more unproductive and the basic institutional structure even less conducive to productive activity. Such a path can persist because the transaction costs of the political and economic markets of those economies together with the subjective models of the actors do not lead them to move incrementally toward more efficient outcomes.

This study sheds light on these contrasting stories by providing a theoretical foundation to the study of institutional change. The next chapter explores the theoretical foundations of the underlying role of institutions — the problem of human cooperation. Then come two key chapters that provide the basic building blocks of a theory of institutions. In Chapter 3 I explore, critically, the behavioral assumptions we employ and suggest modifications in those behavioral assumptions, and in Chapter 4 I provide a theoretical foundation to the costliness of exchange and its surprisingly important but unappreciated implications.

In the next three chapters I successively describe three dimensions of institutions: formal rules and informal constraints, and the effectiveness of their enforcement. Then I am in the position in Chapter 8 to tie together the threads and illustrate the relationship between institutions and transaction and transformation (production) costs.
Institutions

Part II provides a framework to analyze institutional change. Chapter 9 explores organizations and the way they interact with institutions. Chapter 10 deals with the stability characteristics of institutions, which are essential to understanding the nature of institutional change. The change we observe is seldom discontinuous (although I shall explore revolutionary change) but instead is incremental, and the nature of the incremental institutional change together with the imperfect way by which the actors interpret their environment and make choices accounts for path dependency and makes history relevant (Chapter 11).

Part III relates institutions and the way they change to economic performance. In Chapter 12 I consider the theoretical implications of institutional analysis for the performance of economies both at a moment of time and over time. Chapters 13 and 14 apply the analytical framework to economic history. Chapter 13 explores the characteristics of institutional change of successively more complex economies in history and contrasts the stable forms of historical exchange with the dynamic institutional change of Western Europe that led to modern economic growth. The final chapter suggests the implications of systematically integrating institutional analysis into economic history and presents some extended historical applications.
Cooperation: the theoretical problem

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. It is most striking in economics, where the contrast between the logical implications of neoclassical theory and the performance of economies (however defined and measured) is startling. Certainly neoclassical theory has been a major contribution to knowledge and works well in the analysis of markets in developed countries. At the other end of the scale, however, it does not provide much insight into such organizations as the medieval manor, the Champagne fairs, or the suq (the bazaar market that characterizes much of the Middle East and North Africa). Not only does it not characterize these organizations' exchange process very well, it does not explain the persistence for millennia of what appear to be inefficient forms of exchange.

The disparity in the performance of economies and the persistence of disparate economies through time have not been satisfactorily explained by development economists, despite forty years of immense effort. The simple fact is that the theory employed is not up to the task. The theory is based on the fundamental assumption of scarcity and hence competition; its harmonious implications come from its assumptions about a frictionless exchange process in which property rights are perfectly and costlessly specified and information is likewise costless to acquire. Although the scarcity and hence competition assumption has been robust and has provided the key underpinnings of neoclassical theory, the other assumptions have not survived nearly so well.

For the past thirty years, other economists and other social scientists have been attempting to modify and refine the issues to see just what has been missing from the explanation. Put simply, what has been missing is an understanding of the nature of human coordination and cooperation. Now, that certainly should not surprise a disciple of Adam Smith. Smith was concerned not only with those forms of cooperation that produced...
Institutions
collusive and monopolistic outcomes, but also with those forms of cooperation that would permit realization of the gains from trade. However, the confusion and misunderstanding that followed on the heels of Ronald Coase's famous "The Problem of Social Cost" (1960) makes clear how difficult it is for economists to come to terms with the role of institutions in capturing the potential gains from trade. Coase said a number of fundamentally important things in both this essay and his "The Nature of the Firm" (1937). The most important message, one with profound implications for restructuring economic theory, is that when it is costly to transact, institutions matter. And as Wallis and North (1986) have demonstrated in their measurement of the transaction costs going through the market (the transaction sector) in the U.S. economy, it is costly to transact.

I
If economists have been slow to integrate institutions into their theoretical models, they, along with other social scientists, have been quick to explore the problems of cooperation in a game theoretic framework. To apply this approach briefly and in an oversimplified fashion, wealth-maximizing individuals will usually find it worthwhile to cooperate with other players when the play is repeated, when they possess complete information about the other players' past performances, and when there are small numbers of players. Such a crude summary disguises the richness (and ingenuity) of the results of an army of game theorists who have extended, elaborated, and modified (as well as found exceptions to) each of those qualifications to squeeze a great deal more out of them. In subsequent chapters I shall have more to say about game theory, because it provides an excellent foil (very much like the pure neoclassical economic model) against which to compare actual performance.

Let me turn the game upside down. Cooperation is difficult to sustain when the game is not repeated (or there is an end game), when information on the other players is lacking, and when there are large numbers of players. These polar extremes in fact reflect real life contrasts. We usually observe cooperative behavior when individuals repeatedly interact, when they have a great deal of information about each other, and when small numbers characterize the group. But at the other extreme, realizing the economic potential of the gains from trade in a high technology world of enormous specialization and division of labor characterized by impersonal exchange is extremely rare, because one does not necessarily have repeated dealings, nor know the other party, nor deal with a small number of other people. In fact, the essence of impersonal exchange is the antithesis of the condition for game theoretic cooperation. But the modern Western world does in fact exist. How come? A neat, definitive an-
Cooperation: the theoretical problem

The noncoincidence of wealth-maximizing behavior and socially cooperative outcomes has been a key factor in the way game theory has evolved. The so-called prisoner's dilemma that has been a mainstay of game theory is closely allied to Mancur Olson's (1965) free-rider dilemma. Both suggest a discouraging perspective on the problems of human cooperation and coordination. However, the most dismal aspects of Olson's analysis and prisoner dilemma problems reflect the static nature of the analysis and the fact that it is a one-shot game. That is, when the prisoner's dilemma game is played only once, it is a dominant strategy for players to defect and therefore not to achieve what would be an efficient outcome with respect to the aggregate well-being of the players. However, it is well known that defection is not necessarily the dominant strategy if the situation is repeated over and over again, as many collective action problems are. In an iterated prisoner's dilemma game, one that is repeated, there is no dominant strategy. In a now-famous tournament, Robert Axelrod found that the winning strategy under these conditions of continuous repeated play is a strategy of tit-for-tat, one in which a player responds in kind to the action of the other player. This led to Axelrod's celebrated *The Evolution of Cooperation* (1984), an optimistic book about the ability of human beings to devise cooperative solutions to problems without the intervention of a coercive state.

The conditions under which cooperation can be sustained have produced an immense literature, both in game theory and by nongame theorists who are interested in the political-modeling process. Three works that focus on the issues and problems of the maintenance of cooperation will, I believe, highlight the issues we are concerned with in this study.

Russell Hardin (1982) focuses on the *n*-person prisoner's dilemma (PD) and explores the difficulties of collective action in large groups. Hardin emphasizes that the difficulties of collective action depend not just on the size of the group, but also on the ratio of costs to benefits. Conventions (which lead to some form of social order) may arise, particularly when there are asymmetries through which the participants may explore each other's motivations and capabilities in iterated games. Hardin argues that

1In a recent historical study of the formation of property rights in natural resource industries in the United States, Libecap (1989) comes to a similar conclusion with respect to the critical role of the ratio of benefits to costs as a determinant of success in efficient property rights formation.
Institutions

conventions may also come into existence when the participants adopt conditional strategies. However, the conditional strategies involve policing and enforcement (by threats).

Michael Taylor (1982, 1987) explores the conditions under which social order can be maintained in anarchy, that is, without the state. He asserts that community is essential for anarchic social order and that the key features of community are shared common beliefs or norms, direct and complex relationships between members, and reciprocity. Taylor argues that the state destroys the very elements of community (an argument that has been made by Titmuss and others) and indeed, to the degree that altruism plays a role, it too can be minimized or destroyed by the coercive action of the state.

Howard Margolis (1982) develops a model in which individual behavior is in part determined by altruistic motives. Margolis argues that individuals have two types of utility functions, those that favor group-oriented preferences and those that favor selfish preferences, and that individuals make trade-offs between the two. His model allows him to explain certain patterns of voting behavior that do not appear to make sense in the behavioral context of a wealth-maximizing individual.

These three works represent major efforts to examine the conditions by which cooperation can be maintained. It is important at this point to confront an issue that will be a focus of this study: that is, under what conditions can voluntary cooperation exist without the Hobbesian solution of the imposition of a coercive state to create cooperative solutions? Historically the growth of economies has occurred within the institutional framework of well-developed coercive polities. We do not observe political anarchy in high-income countries. On the other hand the coercive power of the state has been employed throughout most of history in ways that have been inimicable to economic growth (North, 1981, Chapter 3). But it is difficult to sustain complex exchange without a third party to enforce agreements. Surely, the jury is still out on what continues to be the fundamental issue for the solution of problems of humankind. Perhaps the most pessimistic perspective is that the arguments of Michael Taylor on community and cooperative solutions do not appear to be viable with large numbers and incomplete information. Norman Schofield, in a perceptive review article on these three works, describes the problem as follows:

The fundamental theoretical problem underlying the question of cooperation is the manner by which individuals attain knowledge of each others preferences and likely behavior. Moreover, the problem is one of common knowledge, since each individual, \( i \), is required not only to have information about others preferences, but also to know that the others have knowledge about \( i \)'s own preferences and strategies.
Cooperation: the theoretical problem

In the restricted N-person PD, it might be possible to argue that this problem is partially resolvable, in the sense that certain types of actors might have good reason to believe that others are of a particular type. In the restricted context of a community, Taylor's argument makes good sense: social norms will be well understood and will provide the basis for common knowledge and this knowledge will be maintained by mechanisms designed to make acts intelligible. In more general social situations, however, individuals will be less able to make reasonable guesses about other individuals' beliefs. The theoretical problems underlying cooperation can be stated thus: what is the minimal amount that one agent must know in a given milieu about the beliefs and wants of other agents to be able to form coherent notions about their behavior and for this knowledge to be communicable to the others? It seems to me that this problem is the heart of any analysis of community, convention, and cooperation. (Schofield, 1985, pp. 12-13)

II

Game theory highlights the problems of cooperation and explores specific strategies that alter the payoffs to the players. But there is a vast gap between the relatively clean, precise, and simple world of game theory and the complex, imprecise, and fumbling way by which human beings have gone about structuring human interaction. Moreover, game theoretic models, like neoclassical models, assume wealth-maximizing players. But as some of the experimental economics literature demonstrates, human behavior is clearly more complicated than can be encompassed in such a simple behavioral assumption. Although game theory demonstrates the gains from cooperating and defecting in various contexts, it does not provide us with a theory of the underlying costs of transacting and how those costs are altered by different institutional structures. It is necessary to return to the Coase theorem to sort out those issues.

Coase began his essay (1960) by arguing that when it is costless to transact, the efficient competitive solution of neoclassical economics obtains. It does so because the competitive structure of efficient markets leads the parties to arrive costlessly at the solution that maximizes aggregate income regardless of the initial institutional arrangements. The arrangements can be circumvented or even changed in a setting of costless transacting. Now to the extent that these conditions are mimicked in the real world, it is because competition is strong enough via arbitrage and efficient information feedback to approximate the Coase zero transaction cost conditions and the parties can realize the gains from trade inherent in the neoclassical argument. That is, competition eliminates the incomplete and asymmetric information that rewards defection in the game theory models.

But the informational and institutional requirements necessary to achieve these results are stringent. They entail that the players not only
Institutions

have objectives, but choose the correct way to achieve them. But how do the players know the correct way (that is, have the correct theory that will allow them) to achieve their objectives? The neoclassical answer, embodied in substantive (or instrumental) rationality models, is that even though the actors may initially have diverse and erroneous models, the informational feedback process (and arbitraging actors) will correct initially incorrect models, punish deviant behavior, and lead surviving players to the correct models.

An even more stringent implicit requirement of the discipline-of-the-competitive-market model is that when there are significant transaction costs, the consequent institutions of the market will be designed to induce the actors to acquire the essential information that will lead them to the correct models. The implication is not only that institutions are designed to achieve efficient outcomes, but that they can be ignored in economic analysis because they play no independent role in economic performance.

None of these stringent requirements can survive critical scrutiny. Individuals act on incomplete information and with subjectively derived models that are frequently erroneous; the information feedback is typically insufficient to correct these subjective models. Institutions are not necessarily or even usually created to be socially efficient; rather they, or at least the formal rules, are created to serve the interests of those with the bargaining power to devise new rules. In a zero-transaction-cost world, bargaining strength does not affect the efficiency of outcomes, but in a world of positive transaction costs it does and given the lumpiness of institutions, it shapes the direction of long-run economic change.

If economies realize the gains from trade by creating relatively efficient institutions, it is because under certain circumstances the private objectives of those with the bargaining strength to alter institutions produce institutional solutions that turn out to be or evolve into socially efficient ones. The subjective models of the actors, the effectiveness of the institutions at reducing transaction costs, and the degree to which the institutions are malleable and respond to changing preferences and relative prices determine those circumstances. Therefore, we next explore the underlying determinants of human behavior, the costs of transacting, and the makeup of institutions.
But how do
ory that will
wer, embod-
is that even
models, the
 correct ini-
rviving play-
pline-of-the-
 transaction
ed to induce
them to the
are designed
in economic
performance.
trutiny. Indi-
derived mod-
is typically
re not neces-
y, or at least
with the bar-
world, bar-
it in a world
divisibilities
in economic
ively efficient
ivate objec-
s produce
ally efficient
the institu-
the institu-
and relative
explore the
acting, and

3

The behavioral assumptions in a theory of
institutions

All theorizing in the social sciences builds, implicitly or explicitly, upon
conceptions of human behavior. Some of the approaches rest on the ex-
pected-utility assumption in economic theory or the extension of that
behavioral assumption into other social science disciplines, loosely
termed rational choice theory. Other approaches raise some quite funda-
mental questions about the traditional economic approach. Although I
know of very few economists who really believe that the behavioral as-
sumptions of economics accurately reflect human behavior, they do
(mostly) believe that such assumptions are useful for building models of
market behavior in economics and, though less useful, are still the best
game in town for studying politics and the other social sciences.
I believe that these traditional behavioral assumptions have prevented
economists from coming to grips with some very fundamental issues and
that a modification of these assumptions is essential to further progress in
the social sciences. The motivation of the actors is more complicated (and
their preferences less stable) than assumed in received theory. More con-
troversial (and less understood) among the behavioral assumptions, usu-
ally, is the implicit one that the actors possess cognitive systems that
provide true models of the worlds about which they make choices or, at
the very least, that the actors receive information that leads to con-
vergence of divergent initial models. This is patently wrong for most of
the interesting problems with which we are concerned. Individuals make
choices based on subjectively derived models that diverge among indi-
viduals and the information the actors receive is so incomplete that in
most cases these divergent subjective models show no tendency to con-
verge. Only when we understand these modifications in the behavior of
the actors can we make sense out of the existence and structure of institu-
tions and explain the direction of institutional change. In this chapter I
first examine expected utility theory, then explore issues of motivation
and the relationship between the complexity of the environment and the
I

What behavior then is consistent with an institution-free world (or at least one where the institutions function costlessly)? I begin by quoting Mark Machina's characterization of what is meant by expected utility theory, which is the underlying behavioral assumption of neoclassical economics:

As a theory of individual behavior, the expected utility model shares many of the underlying assumptions of standard consumer theory. In each case we assume that, the objects of choice, either commodity bundles or lotteries, can be unambiguously and objectively described, and that situations which ultimately imply the same set of availabilities (e.g., the same budget set) will lead to the same choice. In each case we also assume that the individual is able to perform the mathematical operations necessary to actually determine the set of availabilities, e.g., to add up the quantities in different size containers or calculate the probabilities of compound or conditional events. Finally, in each case we assume that preferences are transitive, so that if an individual prefers one object (either a commodity bundle or a risky prospect) to a second, and prefers this second object to a third, he or she will prefer the first object to the third. (Machina, 1987, pp. 124-5)

In the past twenty years, this approach has come under severe attack and also has found strong defenders. The severe attack has come from experimental economic methods, research by psychologists, and other empirical work, all of which have revealed major empirical anomalies associated with this approach.1 Briefly, these fall into the following categories: violations of the transitivity assumptions; framing effects, where alternative means of representing the same choice problem can yield different choices; preference reversals, where the ordering of objects on the basis of their reported valuations contradicts the ordering implied in direct choice situations; and problems in the formulation, manipulation, and processing of subjective probabilities in uncertain choices.

Most of these anomalies have emerged in the context of carefully date

1The extensive literature dealing with these issues is best seen in the proceedings of a conference held at the University of Chicago in October 1985 entitled The Behavioral Foundations of Economic Theory (Hogarth and Reder, eds.). At this conference a large number of psychologists, economists, and a few members of other social science disciplines gathered and explored fruitfully the complexities and issues involved in the behavioral analysis employed by economists. In addition, see the survey by Mark Machina in the first issue of the Journal of Economic Perspectives (1987), the 1987 Annual Lecture to the Scottish Economic Society given by Frank Hahn (Hahn, 1987), and Rationality in Economics by Shaun Hargreaves-Heap (1989).

2With...
signed experiments, which deal with rather limited sets of issues. As I shall be at pains to discuss later in this chapter, they do not appear directly applicable to the immediate subject here, which is the role of behavioral assumptions in the formation and indeed in the existence of institutions. But they do form the basis for thinking critically about the set of issues we must examine.

Perhaps the best summary of the neoclassical behavioral assumptions was made by Sidney Winter. He argues that there are seven steps to what he calls the classic defense of neoclassical behavioral assumptions. They are:

1. The economic world is reasonably viewed as being in equilibrium.
2. Individual economic actors repeatedly face the same choice situations or a sequence of very similar choices.
3. The actors have stable preferences and thus evaluate the outcomes of individual choices according to stable criteria.
4. Given repeated exposure, any individual actor could identify and would seize any available opportunity for improving outcomes and, in the case of business firms, would do so on the pain of being eliminated by competition.
5. Hence no equilibrium can arise in which individual actors fail to maximize their preferences.
6. Because the world is in approximate equilibrium, it exhibits at least approximately the patterns employed by the assumptions that the actors are maximizing.
7. The details of the adaptive process are complex and probably actor and situation specific. By contrast, the regularities associated with optimization equilibrium are comparatively simple; considerations of parsimony, therefore, dictate that the way to progress in economic understanding is to explore these regularities theoretically and to compare the results with other observations.²

It is important to emphasize a particular point here. The behavioral assumptions that economists use do not imply that everybody's behavior is consistent with rational choice. But they do rest fundamentally on the assumption that competitive forces will see that those who behave in a rational manner, as described above, will survive, and those who do not will fail; and that therefore in an evolutionary, competitive situation (one that employs the basic assumption of all neoclassical economics of scarcity and competition), the behavior that will be continuously observed will be that of people who have acted according to such standards. Before I criticize this argument and its extension to institutional economic theo-

Institutions

ry, it is important to note very carefully its successes. In those instances where something approximating the conditions described above exist, the neoclassical model has been a very effective model for analyzing economic phenomena. For example, in the study of finance, where financial markets tend to have many of the characteristics described above, substantial successes have been made using the straightforward assumptions just described.3

II

To explore the deficiencies of the rational choice approach as it relates to institutions, we must delve into two particular aspects of human behavior: (1) motivation and (2) deciphering the environment. Human behavior appears to be more complex than that embodied in the individual utility function of economists' models. 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 preexisting mental constructs through which they understand the environment and solve the problems they confront. Both the computational abilities of the players and the complexity of the problems to be solved must be taken into account in understanding the issues. We explore first the motivation of the actors.

In recent years the work of sociobiologists and economists has been combined to explore the many parallels between the underlying features of genetic survival and evolutionary development among animals and similar patterns of behavior among human beings. Many economists have found that this approach is not only congenial, but that it also reveals a great deal about human behavior. Jack Hirshleifer (1987) compares biological evolutionary models with socioeconomic ones as follows:

Evolutionary models share certain properties. First of all, they concern populations. Even where we seem to be speaking of single entities, if the course of change is evolutionary it can be described in terms of changing populations of micro-units. Thus, the evolutionary course of a disease within a single human body is a function of the relations among populations of bacteria, antibodies, cells, and so on. Or the evolution of a single nation's economy is the result of changing relations among populations of individuals, trading units, and the like. Evolutionary models represent a combination of constancy (inheritance) and variation. There must be an unchanging as well as a changing element, and even the changing

3The essays by Charles Plott and Robert Lucas in Hogarth and Reder (1986) provide a thoughtful defense of the assumptions of the neoclassical model in specific contexts.
Behavioral assumptions

element itself must be heritable if a system can be said to evolve. In biological
evolution, the emphasis is upon differential survival and reproduction of organ-
ismic types or characters from one generation to the next. Here the constancy is
due to Mendelian inheritance of permanent patterns of coded genetic instructions
(genes). Variation stems from a number of forces, including internal mutations of
these instructions (genetic copying errors), recombination of genes in sexual re-
production, and the external pressure of natural selection. Socioeconomic evolu-
tion mainly concerns the differential growth and survival of patterns of social
organization. The main inheritance element is the deadweight of social inertia,
supported by intentionally taught tradition. As for variation, there are analogues
to mutations (copying errors as we learn traditions). Also, natural selection is still
effective. Finally, imitation and rational thought constitute additional non-genetic
sources of socioeconomic variation. (Hirshleifer, 1987, p. 221)

Efficiency in this evolutionary model does not necessarily have the nice
properties that economists give the term, but frequently is associated with
group dominance at the expense of others. But it also should be noted that
altruism can be a part of the model, as Dawkins has convincingly shown.4
This approach is even consistent with ways by which reputation, trust,
and other aspects of human behavior that on the surface appear to be
altruistic and not consistent with individual wealth-maximization turn
out to be superior survival traits under certain circumstances.5

Thus, we can build more elaborate models of complex human behavior
within the individual expected-utility model, incorporating certain as-
pects of altruism. However an alternative approach, illustrated in Becker’s
study of the family (1981), explores altruism as still another facet of utility
maximization, in which we get utility from the well-being of others. But
this issue is deeper than family altruism. Both research in experimental
economics and a number of studies by psychologists point out that issues
of free-riding, fairness, and justice enter the utility function and do not
necessarily fit neatly with the maximizing postulates in the narrow sense
just described.6 These issues appear to show in the voting behavior of
legislators; it is widely observed that one cannot explain the voting behav-
ior of legislators within the narrow confines of a principal/agent model, in
which the agent (the legislator) is faithfully pursuing the interests of the
principal (the constituents). The agent’s own utility function – his or her
own sense of the way the world ought to be – appears to play a role in the
outcomes.

5See, for example, R. Frank, “If Homo Economicus Could Choose His Own Utility
6See in particular the essay by Kahneman, Knetsch, and Thaler, “Fairness and the
Assumptions of Economics” (1986); Richard Herrnstein, “A Behavioral Alternative
to Utility Maximization” (1988), and Hoffman and Spitzer, “Entitlements, Rights
and Fairness: Some Experimental Results” (1985).
Institutions

The evidence we have with respect to ideologies, altruism, and self-imposed standards of conduct suggests that the trade-off between wealth and these other values is a negatively sloped function. That is, where the price to individuals of being able to express their own values and interests is low, they will loom large in the choices made; but where the price one pays for expressing one's own ideology, or norms, or preferences is extremely high, they will account much less for human behavior (Nelson and Silberberg, 1987). I shall come back to this point, because it helps us to understand a great deal, both about institutions and about the way in which they influence decision making. I intend to demonstrate that institutions basically alter the price individuals pay and hence lead to ideas, ideologies, and dogmas frequently playing a major role in the choices individuals make.

III

The second crucial element in our understanding human behavior is deciphering the environment. This issue plays little or no role in the standard economist’s repertoire, although Lucas (1986) acknowledges that one does not get the consequences of rational expectation models without learning on the part of the players and indeed without the implication of stable equilibria and competition (the implication Winter derives), so that the choices and the alternatives become clearly known. On the face of it, the assumptions of stable equilibrium and knowledge about alternatives are quite attractive, because our lives are made up of routines in which the matter of choices appears to be regular, repetitive, and clearly evident, so that 90 percent of our actions in a day do not require much reflection. But in fact, it is the existence of an imbedded set of institutions that has made it possible for us not to have to think about problems or to make such choices. We take them for granted, because the structure of exchange has been institutionalized in such a way as to reduce uncertainty. As soon as we move away from choices involving personal and repetitive actions to making choices involving impersonal and non-repetitive exchanges the uncertainty about outcomes increases. The more complex and unique the issues we confront, the more uncertain the outcome. We simply do not possess theories to predict effectively the outcomes, and the information we receive in such circumstances frequently does not permit us to update our models to improve them. Herbert Simon has put the issues very well:

If we accept values as given and consistent, if we postulate an objective description of the world as it really is, and if we assume that the decisionmaker’s computational powers are unlimited, then two important consequences follow. First,
we do not need to distinguish between the real world and the decisionmaker's perception of it: He or she perceives the world as it really is. Second, we can predict the choices that will be made by a rational decisionmaker entirely from our knowledge of the real world and without a knowledge of the decisionmaker's perceptions or modes of calculation. (We do, of course, have to know his or her utility function.)

If, on the other hand, we accept the proposition that both the knowledge and the computational power of the decisionmaker are severely limited, then we must distinguish between the real world and the actor's perception of it and reasoning about it. That is to say, we must construct a theory (and test it empirically) of the processes of decision. Our theory must include not only the reasoning processes but also the processes that generate the actor's subjective representation of the decision problem, his or her frame.

The rational person in neo-classical economies always reaches the decision that is objectively, or substantively, best in terms of the given utility function. The rational person of cognitive psychology goes about making his or her decisions in a way that is procedurally reasonable in the light of the available knowledge and means of computation. (Simon, 1986, pp. 5210-11)

Simon's statement captures the essence of why, in my view, the subjective and incomplete processing of information plays a critical role in decision making. It accounts for ideology, based upon subjective perceptions of reality, playing a major part in human beings' choices. It brings into play the complexity and incompleteness of our information and the fumbling efforts we make to decipher it. It focuses on the need to develop regularized patterns of human interaction in the face of such complexities, and it suggests that these regularized interactions we call institutions may be very inadequate or very far from optimal in any sense of the term. In short, such a way of looking at how human beings proceed is consistent with the arguments about the formation of institutions, which I shall discuss later in this chapter.

In "The Origins of Predictable Behavior" (1983), Ronald Heiner makes many of the same points. He argues that the gap between the competence of the agent in deciphering problems and the difficulty in selecting the most preferred alternatives, what he calls the CD gap, is a major key to the way in which human beings behave. His essay is based upon the simple notion that the greater that gap, the more likely the agents will impose regularized and very limited patterns of response to be able to deal with the complexities and uncertainties associated with that gap. Heiner argues, indeed, that this uncertainty not only produces predictable behavior
Institutions

but is the underlying source of institutions. Heiner's essay is unique in its attempt to connect uncertainty and behavior with the creation of institutions. His framework is evolutionary, however, and leaves no room for subjective perceptions of fairness to enter into the behavioral decisions of individuals.

IV

We can summarize the issues discussed above by returning to the classic defense and reacting to the seven points Winter sets forth.

1. For some purposes the concept of equilibrium is a valuable tool of analysis, but for most of the issues that we are concerned with there is not one equilibrium, but multiple equilibria that arise because "there is a continuum of theories that agents can hold and act on without ever encountering events which lead them to change their theories" (Hahn, 1987, p. 324).

2. Although individual actors face many repetitious situations and, as noted above, can act rationally in such situations, they also are confronted with many unique and nonrepetitive choices where the information is incomplete and where outcomes are uncertain.

3. Although Becker and Stigler have made an impressive case (1977) for relative price changes accounting for many apparent changes in preferences, the stability issue is not so easily dismissed. Not only do anomalies show up at the disaggregated level at which psychological research has been conducted, but certainly historical evidence suggests that preferences over time change. I know of no way to explain the demise of slavery in the nineteenth century that does not take into account the changing perception of the legitimacy of one person owning another.

4. Actors would certainly like to improve outcomes, but the information feedback may be so poor that the actor cannot identify better alternatives.

5. Competition may be so muted and the signals so confused that adjustment may be slow or misguided and the classic evolutionary consequences may not obtain for very long periods of time.

6. The condition of the world throughout history provides overwhelming evidence of much more than simple rational noncooperative behavior.

7. The behavioral assumptions of economists are useful for solving certain problems. They are inadequate to deal with many issues confronting social scientists and are the fundamental stumbling block preventing an understanding of the existence, formation, and evolution of institutions.
Behavioral assumptions

It would be nice to conclude this chapter with a precise and tidy behavioral model that not only explained why institutions are a necessary extension of the way human beings process information, but also predicted the complex mix of motivations that shape choices. We have made progress toward doing so; indeed enough to explain the existence of institutions and (less precisely) the motivation of the actors that helps to shape institutions and provides the means by which altruism and other nonwealth-maximizing values enter the choice set.

Institutions exist to reduce the uncertainties involved in human interaction. These uncertainties arise as a consequence of both the complexity of the problems to be solved and the problem-solving software (to use a computer analogy) possessed by the individual. There is nothing in the above statement that implies that the institutions are efficient.

The complexity of the environment is the subject of the next chapter. It is sufficient to say here that the uncertainties arise from incomplete information with respect to the behavior of other individuals in the process of human interaction. The computational limitations of the individual are determined by the capacity of the mind to process, organize, and utilize information. From this capacity taken in conjunction with the uncertainties involved in deciphering the environment, rules and procedures evolve to simplify the process. The consequent institutional framework, by structuring human interaction, limits the choice set of the actors.

There can be no question that the mind's ability to process information is limited, but how does the motivation of the actor enter into the decision-making process? In a strict sociobiological model, maximizing survival potential motivates the actor. Such motivation sometimes, but not always, coincides with wealth-maximizing behavior. The complexity of the environment, given the limited processing ability of the actor, can explain the subjective perceptions of reality that characterize human understanding and even the sense of fairness or unfairness that the individual feels about the institutional environment. To take classic illustrations it is not hard to understand how an industrial proletarian could feel that he or she was being exploited by the bourgeoisie, or how the late-nineteenth-century U.S. farmer could feel the railroad was responsible for his plight. In both cases there were ready-made ideological constructs that explained and accounted for their plight. But the fact that individuals acted upon those perceptions to overcome the free-rider problem is more difficult to explain.

The broad range of human actions characterized by such activities as the anonymous free donation of blood, the dedication to ideological causes such as communism, the deep commitment to religious precepts,
Institutions

or even the sacrificing of one's life for abstract causes could all be dismissed (as many neoclassical economists dismiss them) if they were isolated events. But obviously they are not and they must be taken into account if we are to advance our understanding of human behavior. If our understanding of motivation is very incomplete, we can still take an important forward step by taking explicit account of the way institutions alter the price paid for one's convictions and hence play a critical role in the extent to which nonwealth-maximizing motivations influence choices. We will take such account in succeeding chapters. But first we must examine in detail what it is about the environment that is so complex.