1

Economics With and Without the Law

1.1 The Need for Economic Governance

Most economic activities and interactions share several properties that together create the need for an institutional infrastructure of governance. First, these activities and interactions are opportunities to create or add value. This includes creation of tangible or intangible property (such as improved land, physical and human capital, reputation, and goodwill) and exchange of goods and services. Second, the activities require input from several individuals. We have known since the birth of economics how division of labor enhances productivity; more recently, we have recognized the importance of creation and preservation of common property resources. Third, the interactions are based on explicit or implicit contracts voluntarily made by all the parties involved; exchanges of a good or service for another good or service or money are the main instances of this. Some actions may be unilaterally undertaken by one party but create costs or benefits for another; examples include accidental or deliberate damage to property, or its opposite, namely positive spillovers that improve another's property.

Everyone can potentially benefit from the creation or addition of economic value. However, each participant in the process usually has available to him various actions that increase his own gain, while lowering the others' gain by a greater amount. The only exceptions are situations involving simultaneous exchange of goods or services of immediately verifiable attributes and qualities, but these are a small subset of all economic interactions. In most situations, the participants have opportunities to supply defective goods, shirk on the job, renege on payment, and so on. Williamson (1979, 1985) has coined the term "opportunism" for this whole class of actions that tempt individuals but hurt the group as a whole.

Problems also arise with property rights. If no mechanisms—governmental or non-governmental—exist to deter theft, then any one person can wait for someone else to create property or produce output and then steal it; this usually takes less effort than creating the property or the product oneself. Some may even extort money from others by making threats of destroying their property. Hirshleifer (2001), Grossman and Kim (1985), and others have focused on these problems.

Anticipation of opportunism, theft, or extortion constitutes a strong disincentive to making potentially valuable investments or entering into mutually beneficial contracts in the first place. Therefore if market economies are to succeed, they need a foundation of mechanisms to deter such privately profitable but socially dysfunctional behaviors, and thereby to sustain adequate incentives to invest, produce, and exchange. In other words, markets need the underpinning of institutions of economic governance.

1.2 Economics Taking the Law for Granted

Economists have always recognized the need for governance. However, until relatively recently they assumed that the government, specifically the institution and machinery of the state's law, provided the needed governance. Criminal law, while it has major non-economic functions, also serves to deter theft and some forms of economic fraud. Civil law has economic aspects centrally in its concerns. Contract law can be said to be mainly for the governance of economic activity; laws of tort and liability pertain to contracts as well as non-contractual relationships, both mainly in the economic sphere.

Even the most libertarian economists, who deny the government any useful role in most aspects of the economy, allow that making and enforcing laws that give clear definitions of property rights, and ensuring adherence to voluntary private contracts, are legitimate and indeed essential functions of government, in addition to national defense. Friedman (1962, p. 2) puts this succinctly:

> [The government's] major functions must be to protect our freedom both from the enemies outside our gates and from our fellow-citizens: to preserve law and order, to enforce private contracts, to foster competitive markets.

There seems universal agreement in traditional economics that the framework of law is a necessary condition for a market economy to succeed.

Extreme libertarians regard the government's legal framework as not only necessary but also sufficient for markets to function well—the Coase Theorem (Coase 1960, 1988) says that if property rights are well defined, voluntary contracts can achieve all available economic benefits, including the internalization of externalities and the provision of public goods. Most other economists recognize the possibility of market failure even under the aegis of a well-functioning state law; for them, institutions of law are not a sufficient condition to ensure optimal outcomes from markets. For example, McMillan (2002, pp. ix, x) lists three other elements for markets to work well: "information flows smoothly, . . . , side-effects on third parties are curtailed, and competition is fostered." The opposites of these, namely imperfect information, externalities, and imperfect competition, are well-recognized causes

of market failure, and they can exist regardless of whether a government adequately protects property rights and enforces contracts.

Thus conventional economic theory does not underestimate the importance of law; rather, the problem is that it takes the existence of a well-functioning institution of state law for granted. It assumes that the state has a monopoly over the use of coercion, and that the state designs and enforces laws with the objective of maximizing social welfare. Moreover, until the last 30 years or so, that is, until economics recognized the ubiquity and importance of information asymmetries and transaction costs, the usual implicit assumption was that the law operated costlessly.[1]

This simple view of the law made it possible to achieve faster progress in the research on the economic forces of supply and demand, and of their equilibration in markets; therefore it was a useful abstraction in its time. However, its shortcomings soon hinder rather than help the economic analysis of markets and limit its usefulness. Only advanced countries in recent times come anywhere near the economist's ideal picture, in which the government supplies legal institutions that are guided solely by concern for social welfare and operate at low cost. In all countries through much of their history, the apparatus of state law was very costly, slow, unreliable, biased, corrupt, weak, or simply absent. In most countries this situation still prevails. Markets with such weak underpinnings of law differ greatly from those depicted in conventional economic theory.

Deficiencies of the law are most acute in less-developed countries (LDCs) and in transition economies. For example, Bearak (2000) reports that there are 25 million cases pending before the courts in India, and even if no new ones are filed, it will take 324 years to clear the backlog. Murrell (1996, p. 34) found that laws in many transition economies were "a facade without a foundation." Recent assessments of the effectiveness of the legal system in post-Soviet Russia differ among Western observers, but a fair assessment is that while the *Arbitrazh* court system created to handle commercial disputes has begun to function reasonably well in handing down verdicts, getting these verdicts enforced remains highly problematic, especially for smaller enterprises (Hendley and Murrell 2001; Hay and Shleifer 1998). McMillan and Woodruff (1999, 2000) and others have found similar situations in other transition economies in Eastern Europe and in Vietnam. Djankov and Murrell (2002) survey this literature.

Of course economic activity does not grind to a halt because the government cannot or does not provide an adequate underpinning of law. Too much potential value would go unrealized; therefore groups and societies have much to gain if they can create alternative institutions to provide the necessary economic governance.

[1] Barzel (2002, especially Chapters 1 and 2) criticizes the orthodox economic view of the state as a benevolent and costless monopolist in coercion.

They attempt to develop, and sometimes succeed in developing, such institutions of varying degrees of effectiveness. These include self-protection or hired professional protection for property rights (see, for example, Hirshleifer 2001; Gambetta 1993), networks of information transmission, and social norms and punishments for contract enforcement (see, for example, Greif 1993, 1994; Milgrom, North, and Weingast 1990; Gambetta 1993). Indeed, an extreme version of the Coase Theorem says that everything works out in the best feasible way (Stigler 1988). If governance is costly, the least-cost method will get chosen from among the available institutions, whether it be state law or a private alternative. In this view, the emergence of a state or government is itself endogenous, and will occur if, and only if, it is the most efficient mode of governance. But even without going that far, we can recognize that societies will attempt to evolve other institutions, albeit imperfect ones, to underpin their economic activity when state law is missing or unusable. In other words, governmental provision of legal institutions is not strictly necessary for achieving reasonably good outcomes from markets.

Rodrik (2003, pp. 10–16) summarizes the lessons of case studies of several countries as follows.

> Institutions that provide dependable property rights, manage conflict, maintain law and order, and align economic incentives with social costs and benefits are the foundation of long-term growth. ... State institutions are not the only ones that matter. Social arrangements can have equally important and lasting consequences. ... Modest changes in institutional arrangements ... can produce large growth payoffs ... [but] the required changes can be highly specific to the context.

This yields some general lessons for policy-makers in LDCs and transition economies who are contemplating market-oriented reforms and privatizations, and for their economic advisers from Western countries and international organizations. They must recognize that markets will not succeed unless they are supported by adequate governance institutions. The processes of creating the institutions and the apparatus of state law, and of improving them to the point where they function well, are slow and costly. But it is not always necessary to create replicas of Western-style state legal institutions from scratch; it may be possible to work with such alternative institutions as are available, and build on them. Of course, to do this we must have a good understanding of how various institutions of governance work, and of how they interact with each other and with an imperfect state law where that exists. My aim in this book is to contribute to the improvement of this understanding.

1.3 "Lawlessness and Economics" in Context

Where does the study of alternative institutions for protection of property rights and enforcement of contracts, to which I have given the eye-catching albeit strictly inaccurate title "Lawlessness and Economics," stand in relation to many closely related fields of inquiry? Such demarcation of fields or subfields is always a difficult question. Reality is usually a complex mixture, whereas theoretical analysis usually proceeds faster and goes deeper by identifying pure conceptual categories. In the present context, all countries have governments of some kind, of varying scope, competence, and benevolence. And in all of them, significant aspects or components of economic activity are conducted without direct reference or recourse to the state's law. Isolating pure categories from this reality is a matter of judgment, and any attempt at a precise definition or delineation leaves significant exceptions or overlaps. But an attempt must be made, however imperfect, so that researchers and students in neighboring areas can fit this book into the context of their own work.

Lawlessness and Economics can be regarded as a subfield within the broad conceptual framework of the New Institutional Economics. This large and varied body of research has built upon pioneering ideas of Coase (1937, 1960, 1988), North (1990), Williamson (1985, 1996), and others; Williamson (2000) has given us a good recent overview and assessment.

North distinguishes between institutions and organizations. For him, institutions are the overarching framework of rules and constraints, formal and informal, that govern interactions among individuals; constitutions and social norms are examples. Organizations are groups of individuals that operate within the general framework of institutions, and implement the rules and norms of the institutions; examples are legislatures, political parties, and universities. Of course there are interactions and feedbacks between institutions and organizations. The rules and constraints imposed by institutions do not eliminate all freedom for organizations to act, and since organizations have members with differing interests and abilities, interesting issues of "the play of the game" at this level must be analyzed. Institutions can then evolve to alter the rules of the game so as to achieve better outcomes from the play at the organizational level. Finally, individuals interact within the frameworks set up by both institutions and organizations, and these transactions have their costs of information, commitment, and so on. North (1990, pp. 92–104) argues that institutions and organizations attempt to economize on transaction costs, but usually fall short of optimality, especially when changing economic and technological conditions require changed or new institutions. He gives two categories of reasons for the long lags and bottlenecks in the process of institutional change: first, resistance by powerful special interests with stakes in the old system; and second, multiple equilibria and historical accidents.

Others use the same terms in slightly different senses and draw somewhat different distinctions. For example, in his pioneering game-theoretic analysis, Schotter (1981, p. 11) defines institutions as

> a regularity in social behavior that is agreed to by all members of society, specifies behavior in specific recurrent situations, and is either self-policed or policed by some external authority.

Thus his institutions specify the strategies that the individuals should choose. That is, they include aspects of the play of the game as well as the rules. Calvert (1995a,b) develops this idea further, and interprets it more explicitly as specifying the equilibrium that is to be played. Sobel (2002, p. 147) similarly says that

> an individual's expectation of the response to his action is often an important part of the institutional environment; that is, the institutional environment also serves to coordinate beliefs and select equilibria.

And Greif (2000) defines institutions as

> a system of social factors—such as rules, beliefs, norms, and organizations—that guide, enable, and constrain the actions of individuals.

Thus he includes organizations as an example of institutions, not a separate category. Moreover, the beliefs in his analysis are beliefs about the strategies that others would choose in off-equilibrium situations, and therefore serve to select an equilibrium; this accords with Calvert's view of institutions. I conclude that North's conceptual distinction between the rules and the play of the game, leading to the distinction between institutions and organizations, serves a useful purpose of focusing our attention on the different functions, but there are many feedbacks between the two categories blurring the distinction.

Williamson (2000) draws finer distinctions with a four-level classification scheme. At the first (highest or most basic) level stand informal institutions, such as religion, social customs and norms. These are slow to change, over the timescale of centuries or millennia. At the second level is the institutional environment, consisting of formal rules, such as constitutions and laws. The timescale of evolution of these is measured in decades. The play of the game occurs at the third level, and this includes the choice of appropriate modes of governance for each type of transaction, or organizations in North's sense, the aim being to economize on transaction costs. The idea that transactions and governance modes are aligned in this way is Williamson's (1996, p. 12) famous discriminating alignment hypothesis. Finally, the fourth and lowest level contains routine economic activities such as production, employment, market equilibration.

Williamson places other subfields within this scheme. Positive Political Theory (PPT) operates mostly at the second level, focusing on political institutions such as

the executive and the legislature, and their implications for economic performance. Transaction Cost Economics (TCE) is located mostly at the third level. It focuses on economic structures such as firms and on transactions within and across these structures. It studies how the economic structures and transactions respond to the available governance structures and their limitations.

As with all attempts at taxonomy, this leaves some ambiguities and overlaps, and is not universally followed. Most importantly, some other people would locate many informal institutions of social norms at the second rather than the first level. For Williamson, norms and other informal institutions have "mainly spontaneous origins" and "have a lasting grip on the way a society conducts itself" (2000, p. 597). In my opinion this is true of some but not all norms. Societies make conscious efforts to instill some norms into their members, enlisting the help of parents, teachers, media, and leaders of opinion for this purpose. Many of the norms pertain to civic duties such as voting, but others pertain to honesty in economic matters. This process of social conditioning and education can respond to changing needs much faster than the evolutionary timescale. Many of the communities facing collective-action problems that were the subject of Ostrom's (1990) studies created and used such norms. Therefore I will be more flexible in the location of norms in Williamson's scheme. Next, there are important feedbacks from level 3 to level 2, therefore they should be studied jointly.

Where does LLE fit within Williamson's classification? In one sense, it is about what PPT leaves out. LLE asks: if the government's apparatus is absent from the institutional environment of level 2, what takes its place and provides the rules of the game? There is another difference between LLE and PPT. In its relation to economics, PPT is interested in more macro aspects of economic policy and how these are affected by political institutions, for example, how fiscal, monetary, regulatory, and trade policies differ between presidential and parliamentary systems (Persson and Tabellini 2000). The focus of LLE is on the much more microeconomic level of individual transactions. At level 3, LLE shares many concerns with TCE, but it is focused more on the interaction among distinct decision units than on governance within one firm. However, this distinction is not at all clear-cut, as the boundaries of decision units are themselves endogenous and can change in response to changes in transaction technologies and enforcement modes.

Thus the concept of institutions in the New Institutional Economics is somewhat imprecise, and is interpreted differently by different scholars. In my view this is not a serious defect. It is not so important to have a definition that would be valid and rigidly enforced in all contexts; what matters is that we understand what it means in any specific context we are analyzing.

Next I discuss the relationship between LLE and another large and growing field, variously labeled "Institutions and Growth" or "Institutions and Development." This

field studies how the institutions and organizations that make and implement government policy affect economic performance, and attempts to identify what constitutes good governance in this sense. The literature is vast and wide-ranging. Pioneering historical and conceptual perspectives such as North (1990) and Olson (1993) have taught us the importance of securing property rights, especially against the government's own predation. Case studies of De Soto (1989, 2000), the compilation of analytical case studies in Rodrik (2003) (and many other case studies), the theoretical modeling of Grossman (2002) and others, and the combination of theory and empirical observation by Shleifer and Vishny (1998), all reinforce this message.

Construction of quantitative indexes of good governance is an active enterprise; a recent example is Kaufman, Kraay, and Mastruzzi (2003). Much empirical work has established positive correlations across countries between various measures of economic performance and various aspects of good governance such as the rule of law in defining and protecting property rights, accountability of policymakers, transparency of policy, lack of corruption, and simplicity and speed of bureaucratic procedures. Examples are Hall and Jones (1999), Sokoloff and Engerman (2000), La Porta et al (1998, 1999), Rodrik (2000), Acemoglu, Johnson, and Robinson (2001), Acemoglu et al (2002), and Kaufman and Kraay (2002). Of course the positive feedback may indicate reverse causation—the demand for good governance rises with incomes—or common effects of third variables on both. Empirical researchers handle this problem using different instrumental variables that are not affected by current economic performance, for example colonial origin (Hall and Jones), historical measures of morality (Acemoglu, Johnson, and Robinson), and an instrumental variable estimation of reverse causation plus direct knowledge of measurement errors to achieve identification (Kaufman and Kraay). In fact Kaufman and Kraay find a negative reverse causation: across countries, higher incomes on average lead to worse governance. This may be because in a country with higher income, especially if the income derives from natural-resource extraction, there is more rent-seeking and greater likelihood of capture of government and policymaking (see Tornell and Lane 1998). However, Sachs and Warner (2001, p. 836) present evidence to the contrary. Overall, the empirical literature gives good support to the proposition that good governance causes higher incomes and growth.

Lawlessness and Economics differs from this literature in two ways, but each of them also points out the need for building better bridges between the two. First, LLE is concerned with micro issues of governance of individual transactions, specifically with property rights and contracts, and not directly with the macro issues of economic growth or development. One can think of LLE as a potential microfoundation for theories of how governance institutions affect macroeconomic performance. Secondly, LLE in its purest form assumes that the government's law is completely absent or ineffective, and examines purely non-governmental alternatives that attempt to

perform some of the same functions; therefore the quality of government is not directly an issue. In most of the formal models I develop in this book I focus on this part of LLE, and do not examine in any detail the situation where the government itself is predatory or kleptocratic. This is because I believe that a thorough theoretical analysis of the pure case of non-state economic governance constitutes a useful start for further research on the interactions between corrupt or predatory governments and private institutions of governance. In Chapter 2, I do consider interactions between private arrangements that can use better inside information on the one hand, and a state law that must use worse public information on the other. However, the state in that chapter is assumed to be neither corrupt nor predatory. Predation by the state or its agents has been the essence of the problem in many countries and in the literature on the quality of government, and a start at theoretical thinking has been made, for example by Shleifer and Vishny (1998), Hay and Shleifer (1998), and Grossman (1995, 2002). I make one contribution to this line of research: in Section 5.4 I reinterpret standard theories of optimal income taxation to obtain some characterization of the interaction between a predatory government, and citizens trying their best to evade its attempts to take their assets and outputs. Further analysis of predatory governments should be an important part of the agenda for future theoretical research in LLE.

I hope this discussion of what LLE does and does not attempt will help readers locate my topic in relation with other fields and subfields. However, the unavoidable ambiguities of taxonomy require that such boundaries should be permeable. In this chapter I will indulge in a few more attempts at definition and demarcation. But then, for the rest of the book, I will proceed without any consistent attempt to stay strictly within those confines.

1.4 Law and Economics

The field of Law and Economics became established during the 1960s and 1970s, and continues to grow. It studies the interaction between state law and economic activities and outcomes. Thus "Law and Economics" and "Lawlessness and Economics" can be regarded as two mutually exclusive and jointly exhaustive subfields of the larger field of economic governance; of course in reality there are overlaps and omissions. In this section I mention some of the major concerns of Law and Economics to bring out the contrasts, but do not go into any of them in any detail, merely referring interested readers to Posner and Parisi (1997). This three-volume handbook contains many of the important articles on Law and Economics, along with detailed editorial commentary.

The early contributions to Law and Economics mostly dealt with the implications of law for economics, that is, they considered the effect of legal rules on the economic choices of individuals and firms and on market outcomes. To quote from

the survey by Cooter and Rubinfeld (1989, p. 1068), this theory "treats laws, like prices, as incentives for behavior." This work encompassed both criminal and civil laws. Becker's (1968) article on the optimal detection and punishment strategies to deter crime has become classic. Modern theory has added information- and game-theoretic aspects to the analysis. Polinsky and Shavell (2000) survey this literature. Various branches of civil law—liability, tort, contract, property—govern situations where two or more individuals can enter into a contractual relationship, explicit or implicit, as well as ones where one person's actions have spillover effects on others without any voluntary agreement on their side. These legal rules affect the incentives of individuals to take actions, or to refrain from actions, that carry benefits or costs to others, and that in turn affect overall economic outcomes and efficiency. Baird, Gertner, and Picker (1994) discuss many examples of this kind in the course of their exposition of game theory.

Later work took up a reverse causation from economics to law and jurisprudence, namely evaluation of alternative legal rules in the light of the economic concepts of efficiency. This led to issues of design or reform of specific legal rules, and even the concept of the law as a system that can be designed, rather than as a collection of cases and precedents to be studied and applied. Posner (2002) examines the law in this way from an economist's perspective.

1.5 Economics in the Shadow of the Law

Even in modern countries where a well-functioning institution and apparatus of government-provided law exists, economic—or, indeed, non-economic—disputes do not immediately lead to litigation. Recourse to the law is often the last resort, not the first one. People attempt to resolve their disputes using various private methods of negotiation, and only if these fail do they go to the courts. In the context of business, this idea goes back at least as far as Macaulay (1963). Williamson (1996, pp. 10, 122), citing previous legal scholars, says that businessmen "speak of 'cancelling the order' rather than 'breaching our contract'," and that contracts and courts are "a norm of ultimate appeal when the relations cease in fact to work." In matters of personal relationships, too, Mnookin and Kornhauser (1979) cite estimates that less than 10% of divorces are contested in court. All of this has led to the concept of "private ordering in the shadow of the law."

An obvious explanation for the persistence of such private ordering is that resolution of disputes using the formal machinery of state law is far from costless; in fact its costs, especially time costs, often exceed those of alternative methods of private ordering. Sometimes formal law may yield outcomes that are worse for all parties than can private ordering. Therefore the outcome that the parties expect to obtain in the court (net of the costs of using the court system) becomes a backstop or threat

point to private negotiation. I will sketch this idea in somewhat greater detail in Section 2.2.

Long-term relationships and arbitration are the most common modes of private ordering. Long-term relationships can be self-enforcing for reasons familiar from the theory of repeated games: the immediate gains from behaving opportunistically can be offset by future losses, because the opportunism leads to a collapse of the relationship and therefore to lower future payoffs. But now we have another possibility. The relationship need not collapse completely; it can be replaced by one based on formal contracts and court enforcement. This is still costly and therefore serves to deter opportunism in the original ongoing relationship. But the cost may be less than that of a total breakdown of interaction. Correspondingly, the deterrence effect falls short of the level possible when total breakdown is the only alternative. Thus availability of court enforcement may, in a seeming paradox, reduce the extent of good behavior that can be sustained in the long-term relationship. I will examine this possibility in Section 2.3.

The official law can interfere with a long-term relationship in another way, namely by agreeing to hear a case filed by one of the parties in such a relationship attempting to overturn an adverse outcome in the implicit contract. However, courts often recognize the merits of implicit contracting in long-term relationships like employment, and refuse to hear such cases. This is the doctrine of forbearance (Williamson 1996, p. 27).

Private arbitration can have cost advantages over the government's courts, but perhaps more importantly, it can have information advantages and therefore provide dispute resolution of higher quality. Arbitration forums specialize by industry, geographic region, and so on, in the range of disputes they take up. They acquire expertise in their special areas. They can adopt procedures and rules of evidence that suit their specific concerns. State courts must stand ready to consider all matters that could arise under the law, and although some attempt can be made to assign cases to judges on the basis of their expertise, the rules and procedures must remain the same for all cases. For these reasons, arbitrators are better able to obtain, interpret, and use information pertinent to the dispute than are the state courts. I will develop a model of such informational advantage in Section 2.4.

Arbitrators lack the coercive powers of the state and therefore cannot ensure compliance with their verdicts. But, as with long-term relationships, the government's courts often recognize the advantages of arbitration for governance of particular classes of transactions. Then they accept the arbitrator's verdict and will not agree to rehear the issue. This is formalized in US laws (Bernstein 2001, footnote 111) and in international agreements (Mattli 2001, p. 939). Given this shadow of the law, if one party refuses to comply with the arbitrator's verdict, the other can enlist the help of the courts for enforcement.

Landes and Posner (1979) describe and analyze private institutions of adjudication historically as well as in the context of modern states. Milgrom, North, and Weingast (1990) describe a specific historical institution, namely the *lex mercatoria* or merchants' law, usually called "the law merchant," in medieval Europe. They and others find that many principles developed by the private judges or adjudicators of the law merchant were later taken over by the state's law. However, Landes and Posner point out that private adjudicators lack the incentives to supply the "public good" of principles and precedents, so we expect it to be underprovided.

1.6 Other Institutions of Economic Governance

Let us turn to an extreme conceptual situation that is the true realm of Lawlessness and Economics, namely an economy lacking any government-provided legal institutions or organizations for protection of property rights and enforcement of contracts. Such a society needs to develop its own alternative modes of economic governance. Two general types of such institutions and organizations are observed. They are parallel to the long-term relationships and arbitration forums mentioned in the previous section, but here the private ordering must operate unsheltered, without the shadow of the law.

The alternative that has been studied most thoroughly is self-enforcing governance through repeated interaction. If the same parties interact with each other repeatedly, and they value the future sufficiently highly relative to the present, then the prospect of a long-term collapse of the relationship can control the temptation to obtain a short-term gain. This is well understood, both in the theory of repeated games and in practice (Axelrod 1984; Abreu 1986; Abreu, Pearce, and Stacchetti 1990; and several others). But many economic activities require dealing with different partners at different times. Even bilateral relationships may get severed, requiring one or both parties to find a different partner in the future. Therefore we must consider situations where there is little long-term relationship with the same person, but stable membership of a whole large group. Self-governance in such a group requires that if any one person cheats his current partner, the news is conveyed to others in the group who might be the cheater's future partners. This loss of reputation can lead to ostracization, or other actions by the group that have the effect of punishing the cheater on behalf of his current victim. In turn, such reputational considerations can deter opportunistic behavior. For this process to work, the society needs good information networks and credible multilateral punishment strategies.

Both of these conditions can be fulfilled in stable and cohesive groups or networks, which might be defined by business ties (Greif 1993; Bernstein 1992, 2001), ethnicity (Casella and Rauch 2002; Rauch 2001), and so on. However, the quality of information and the credibility of punishment both degrade as the size of such a group increases. The case studies of Ostrom (1990) and the contrasting case studies

of two merchant groups by Greif (1994, 1997) illustrate this vividly, and numerical calculations on the theoretical models of Kandori (1993) and Ellison (1994) show similar tendencies. Therefore we need a better understanding of the limits of self-governance. What happens if trading opportunities expand beyond the close group? When does some other mode of governance become better? What happens at the interface between the two systems? I will tackle some of these questions in Chapter 3.

Instead of relying on self-governance, the group might attempt to obtain the service for a fee from a private individual or group. One can think of this as a "private government," established to serve just this one function, as opposed to the broader institution we call the government, which performs a multitude of functions. Credit-rating agencies and similar certification intermediaries can collect and disseminate information about a person's history to his prospective partner. Arbitration is another common arrangement of this kind. In the absence of state law, it cannot rely on the courts' forbearance to ensure compliance with its verdicts. But arbitration can use repeated interactions in the group. Thus, if any member of the group defies the arbitrator's ruling, the arbitrator can publicize this information to the whole group, and then the group will not deal with the miscreant. In effect, the arbitrator becomes the hub of an information network. The private judges at trade fairs in medieval Europe functioned in this way (Milgrom, North, and Weingast 1990); Bernstein (1992, 2001) gives examples of modern trade associations that provide similar functions for their members.

Finally, organized crime provides services of information as well as enforcement (Gambetta 1993; Varese 2000; Whiting 1999; and others). In a society without state law, there is no external mechanism to ensure honesty of the arbitrator, the private judge, or the mafioso. That has to be self-enforcing, based on reputation considerations in a long-term relationship. Even though the participants in the economic transactions may not meet the same partners repeatedly, each of them can have a repeated interaction with the person or organization that provides the governance, so an honest equilibrium of this kind is logically conceivable. I will study the operation of such third-party private governance, paying attention to the question of the intermediary's honesty, in Chapter 4.

Thus far I have been concerned with the governance of economic interactions, that is, explicit or implicit contracts between two or more parties, made with mutual consent. Protection of property rights raises some different issues. Violation of property rights is a unilateral action taken by the predator; this differs from contractual relationships, which are based on voluntary consent of both or all parties. Indeed, the owner of a property may not even know the identity of the potential thief or extortionist. Then the potential victim must take unilateral steps to deter the potential invader, and to detect and punish him if deterrence fails. The prop-

erty owner may try to do this directly, diverting resources from other productive uses into protection, or he may hire a specialized protector—a private guard or, again, organized crime. In some countries and at some times the government or its agents may be the thieves who try to extract as much as they can from the citizens for their own consumption. The citizens cannot hope to resist the government's coercive power with force, but may attempt to hide their assets. Also, the prospect that the fruit of ones's efforts will be taken by the government will be a disincentive to produce or accumulate. I will examine some of these issues in Chapter 5.

Anecdotal evidence suggests that the various alternative institutions of governance can be very effective. Greif (1993, p. 528) found "only a handful of documents contain[ing] allegations of misconduct" in the archives of the correspondence among Maghribi traders. Bernstein (1992) reports that in the numerous transactions that occur every year among the 2000 members of the New York Diamond Dealers' Club and the numerous non-members who trade there, only 30–40 trades result in a judgment from the arbitration system of the club. Exact figures are not available for the total number of transactions or the number of cases where the defendant refuses to pay the judgment, but a safe guess is that the former is in the hundreds of thousands and the latter in single digits.

Some may regard this as evidence for fundamental goodness of human nature. However, the record of failures of other less-well-designed institutions of governance suggests otherwise. Ostrom (1990, Chapter 5) and Liebcap (1989, Chapters 5, 6) discuss cases of failures to define and enforce property rights and to solve common-resource-pool problems. Several studies in Rodrik (2003) show that less-developed countries with poor property-right and contract-enforcement systems fail to attract foreign investment and sustain growth. Thus we should conclude that institutions can be effective deterrents to opportunism, but that in their absence, beneficial economic activity is likely to be hindered by a well-grounded fear of being cheated.

1.7 Some Basic Analytical Apparatus

The need for governance arises because, in its absence, individuals pursuing their own interests would generate an inferior equilibrium outcome. Game theory studies many instances of this, most notably the prisoner's dilemma. In this section I briefly outline three simple games of this kind; in later chapters, some extensions and variants will be developed and applied in greater detail. Readers who are familiar with the theory of repeated games can omit this section.

First consider a situation involving two participants, or players in the sense of game theory. The first player begins the game by choosing whether to take a costly action, here called an investment. Then the second player takes another action that determines how the product of this investment is shared between the two. They have an agreement or understanding on this matter. If the second player follows through

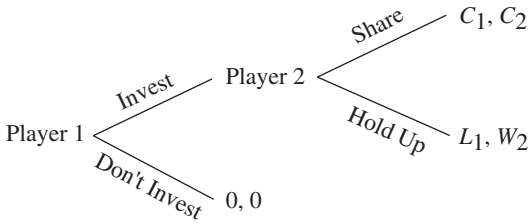

Figure 1.1. One-sided prisoner's dilemma in extensive form.

Table 1.1. One-sided prisoner's dilemma in normal form.

		Player 2	
		Share	Hold Up
Player 1	Invest	C_1, C_2	L_1, W_2
	Don't Invest	0, 0	0, 0

with the agreement, both parties will get a positive outcome or game-theoretic payoff. However, the second player can instead take an opportunistic action, which will yield him a larger payoff but the first player a negative payoff. This is the famous "hold-up problem" (Klein, Crawford, and Alchian 1978; Williamson 1979). It is also the situation of medieval merchants studied by Greif (1993, 1994, 1997a, 2000). These merchants had to consign goods to distant agents, who might then abscond with the goods or the proceeds of the sales.

Figure 1.1 shows the game in its extensive or tree form and Table 1.1 shows the payoff matrix in the strategic or normal form. If Player 1 does not make the investment, the status quo prevails. This is chosen as the origin for measuring the payoffs of both players, so both get 0 in this outcome. If the first player invests and the second takes the sharing action, the resulting positive payoffs are denoted by C_1 and C_2, respectively. If the first player invests and the second engages in opportunistic hold-up, then the first gets a loser's payoff L_1 worse than the status quo ($L_1 < 0$), while the second gets a winner's payoff W_2 better than that from sharing ($W_2 > C_2$).

In the extensive form, the second player's optimal choice at the second node is Hold Up ($W_2 > C_2$), and foreseeing this, the first player's optimal action is Don't Invest ($L_1 < 0$). This is the only outcome of rollback analysis, or subgame-perfect equilibrium. And in the normal form, (Don't Invest, Hold Up) is the only Nash equilibrium; for any other strategy combination, one of the players wants to deviate to a different strategy. In this outcome both players get 0, but both would be better off if they could achieve (Invest, Share). The second player could promise that he will choose Share, but in the absence of some form of governance, the promise is not credible.

This is like a prisoner's dilemma except that only the second player has the opportunity to make an extra private gain, therefore it is often called a one-sided prisoner's dilemma. There are two broad approaches to resolving such dilemmas, although each appears in many different manifestations appropriate for different circumstances.

First, some method can be found for inflicting a direct penalty on the second player if he chooses Deviate. The state law can do this. The two players agree before the fact that (Invest, Share) is the better outcome, so they can sign a formal contract whereby the first player undertakes to choose Invest and the second to respond Share. The court stands ready to enforce this contract, under the threat of sufficiently harsh fines, and ultimately coercion if that proves necessary. Of course this requires that any deviations from the agreement can be proved before the court, or in game-theoretic terminology, are verifiable. If state law is absent or ineffective, private enforcement can perform similar functions to some extent; that possibility was mentioned in Section 1.6 and is the subject of Chapter 4.

Second, if this game is played repeatedly, the prospect that future benefits depend on current performance may suffice to induce the second player to choose Share, and then the first can recognize this and choose Invest. In the strict logic of game theory, repetition must be infinite or at least indefinite, because in any fixed and finite sequence of repetitions, cooperation will unravel from the end. In reality, some cooperation is observed in cases of fixed repetitions or even single play (Camerer 2003, pp. 45, 46, 221–223), but participants in most commercial transactions would be reluctant to rely on this and would insist on the support of some mechanism of governance.

If the same two players continue to interact for the infinite or indefinite future, the good outcome can be sustained as an equilibrium of the repeated game if the interest rate r at which the future is discounted is sufficiently low. Consider the following strategies: Player 1's strategy is "Invest so long as Player 2 has not chosen Hold Up in the past, but stop investing in response to any history of Hold Up." Player 2's strategy is "Share always." Player 1's strategy is called a grim-trigger strategy: it threatens a grim punishment, namely a total breakdown of the relationship, and it is triggered by any action of Player 2 that departs from his specified equilibrium strategy.

For this to be an equilibrium, neither player should stand to gain by deviating from the specified strategy. The key test is for the second player. If he deviates once, he gets W_2 instead of C_2, for an immediate one-time gain of $(W_2 - C_2)$. Then the first player stops investing, so the second player gets 0 for each subsequent period, whereas he would have received C_2 if he had kept to the specified strategy of compliance. If the deviation is to be unprofitable to Player 2, the interest he could earn on the up-front one-time gain should not exceed the future loss in each period.

That is,

$$r\ (W_2 - C_2) \leqslant C_2\,. \tag{1.1}$$

Equivalently, the immediate gain $(W_2 - C_2)$ should not exceed the capitalized value C_2/r of the subsequent per-period loss. This reasoning assumes that the deviating action is immediately detected, and the future cost comes in the form of a permanent collapse of the mutually beneficial arrangement. But the idea can be extended to more complex situations. Detection of a deviation may be less than perfect. In a large population where one player may meet different potential partners in different periods, the news of a deviation may be communicated to others with less than certainty. The population may have people with different preferences or behaviors, and a player's action may convey information about his type. Strategies other than permanent disengagement may be used. However, the general idea of a trade-off between immediate gains and subsequent losses underlies them all. In later chapters I will develop several such applications.

A thematically related but structurally different problem is that of agency. One of the parties to the transaction (the principal) wants the other (the agent) to take some action that will bring the principal some gain but require the agent to expend some cost. The contract specifies how the agent will be compensated for this. But the contract is constrained by information. Some or all of the agent's action and cost and the principal's gain may be observable only to the person in question, or observable to both parties but not provable or verifiable to outsiders, or verifiable to some industry experts but not to the general public including officials of state law. The principal and the agent share a common interest in writing as efficient a contract as the information and institutions allow, but then the agent will behave opportunistically and pursue his own best interest to the extent that the contract allows. This looks like a one-sided dilemma, but it is better handled using different models and techniques. I examine some aspects of such agency problems in Chapter 2.

In many transactions, both sides can engage in opportunism, and their game becomes a conventional (two-sided) prisoner's dilemma. Gambetta (1993, p. 15) gives a nice example. A cattle breeder he interviewed in the course of his research on the Sicilian Mafia told him:

> When the butcher comes to me to buy an animal, he knows that I want to cheat him [by giving him a low-quality animal]. But I know that he wants to cheat me [by reneging on payment].

The game has the familiar payoff matrix shown in Table 1.2. I label the two actions "Comply (with the explicit or implicit terms of the agreement)" and "Deviate (from the agreement)." In informal discussions, I will sometimes refer to these as honest behavior and cheating, respectively, but Comply and Deviate better capture the idea

Table 1.2. Two-sided prisoner's dilemma.

		Player 2	
		Comply	Deviate
Player 1	Comply	C_1, C_2	L_1, W_2
	Deviate	W_1, L_2	D_1, D_2

formally.[2] Each player fares best when he chooses Deviate while the other chooses Comply, and worst when he chooses Comply while the other chooses Deviate. Between these extremes, the payoff is higher when both choose Comply than when both choose Deviate. Thus $W_1 > C_1 > D_1 > L_1$ and $W_2 > C_2 > D_2 > L_2$.

Of course formalization requires simplification of reality. The single action labeled Deviate may in fact stand for a complex set of possibilities. Indeed, the true game of opportunism may consist of seeking and finding actions that are advantageous to oneself, not merely choosing one from a set of already known actions. Another subtlety arises if a formal contract does not constitute the true implicit understanding between the parties, with the result that "Comply" means "comply with the true spirit of the contract," while "Deviate" means "revert to the letter of the formal contract," as in a work-to-rule. Although formalization hides such subtleties, it has its own advantage of giving us a deeper understanding of the mechanisms or processes of the parties' interactions. Therefore formal analysis should be done, and then interpreted in a broader and more flexible manner. That is what I shall attempt.

If the game is played once—or a fixed, finite, and known number of times—each player's dominant strategy is always to play Deviate. The resulting equilibrium with payoffs (D_1, D_2) is worse for both than what they would get if both chose Comply (acted honestly), but that is not achievable given the temptation of each to choose Deviate (cheat).

If the game is repeated infinitely or indefinitely often, the good outcome where both choose Comply can be sustained using grim-trigger strategies for both players, stipulating permanent deviation by both players after the occurrence of even a single deviation by either. When each expects the other to follow this strategy, Player 1 does not want to deviate if the interest $r\ (W_1 - C_1)$ he could earn on his immediate gain $(W_1 - C_1)$ from choosing Deviate does not exceed the reduction in his payoffs for each subsequent period, $(C_1 - D_1)$. That is,

$$r\ (W_1 - C_1) \leqslant C_1 - D_1\ , \tag{1.2}$$

[2]Cooperate and Defect are commonly used in game theory, but I want to avoid "Cooperate" because that invites confusion with a "cooperative game" in the technical sense, where the actions are jointly agreed and also jointly implemented. There the substantive issue is the process of reaching an agreement, and deviations are impossible once an agreement has been reached. By contrast, my focus is on how to achieve good or cooperative *outcomes* despite the fact that the players choose *actions* individually (non-cooperatively).

Table 1.3. Two-sided prisoner's dilemma with option not to play.

		Player 2		
		Play, Comply	Play, Deviate	Don't Play
Player 1	Play, Comply	C_1, C_2	L_1, W_2	0, 0
	Play, Deviate	W_1, L_2	D_1, D_2	0, 0
	Don't Play	0, 0	0, 0	0, 0

and similarly for Player 2. In later chapters I will discuss several variants and elaborations of this game; a general theoretical treatment can be found in Osborne and Rubinstein (1994, Chapter 8), and a survey of the literature in Pearce (1992).

The standard prisoner's dilemma game assumes that both players have already decided to play. However, in many real-life situations the players have the option of not playing, if they calculate that their payoffs in the resulting equilibrium will be less than what they could get in the available alternative opportunities. This is a two-stage game. At the first stage each player decides whether to play the dilemma game. If both choose to play, the second stage occurs and consists of the actual play of the dilemma game. While this could be shown in its extensive form, it is simpler to look at the normal form. Each player chooses one of three strategies: (1) Don't play, (2) Play and choose Comply, and (3) Play and choose Deviate. The payoff scales are chosen so as to make each player's payoff from his outside alternative equal to zero. The other payoffs are as before. Table 1.3 shows the payoff matrix of the game.

The outcome where both players choose "Don't Play" is always a Nash equilibrium of this game, because if one is choosing this action, the other cannot get a higher payoff by choosing anything else. But there can be other equilibria, depending on the signs of D_1 and D_2, that is, whether it is better to play the game even if both are going to Deviate. If D_1 is negative, then there is an equilibrium where Player 1 chooses not to play, and Player 2 chooses Play and Deviate. If D_2 is negative, there is a similar equilibrium with the players interchanged. If both D_1 and D_2 are negative, both of these are equilibria, so there are three equilibria in all. However, since both have to choose to play for the game to take place, the outcome in all these new equilibria is that the game is not played and both players end up with their outside alternative, so these are distinctions without a difference.

If both D_1 and D_2 are positive, we find a genuinely different second Nash equilibrium, where both choose Play and Deviate. The existence of two equilibria is often a nuisance, but it can be a help in sustaining cooperation in repeated games. Specifically, even with a finite and known number of repetitions, cooperation may be sustainable for all but the last few plays. This can be achieved using strategies that specify starting out with the (C_1, C_2) outcome, and switching to the (D_1, D_2) outcome in those last few plays, but switching immediately to the $(0, 0)$ outcome if either player deviates at one of the earlier plays where compliance was indicated.

Benoit and Krishna (1985) develop the general theory of such games; I will have occasion to use a simple specific instance of it in Chapter 3.

The situations studied in each of the following three chapters differ in their specification of players, information, and available strategies. Therefore the models above are not directly applicable merely by a special interpretation of the actions and the payoffs. While they share the schematic structure of the dilemma games, each needs its own reformulation.

All of them have one restrictive feature requiring discussion. All are models of repeated games under stationary conditions. Each can describe how an institution of governance functions, or how two institutions can coexist, once all traders' expectations and actions have adjusted to this fact. However, the process whereby a society changes from one institution to another is dynamic. When I say something about this dynamics, I have interpreted the models beyond their formal results, for example drawing some dynamic inferences from comparative static analysis.

Dynamic games are far more complex than repeated games, and require different approaches and techniques, for example evolutionary theory, stochastic processes, and computer simulations. Some authors, for example Young (1998) and Aoki (2001), have taken this approach to the study of institutions. In comparison with this book, they represent a different choice in the trade-off between a better representation and analytical tractability.

My focus is on the problems of information and enforcement in specific contexts: verifiability and its cost in Chapter 2; localization of communication in Chapter 3; and simultaneous games played between each trader and the intermediary in Chapter 4. To treat these details adequately, I have simplified the dynamics. Others make the opposite choice: they represent the game played each period in a simple or schematic way and focus on the dynamic process. Young (1998, Chapter 9) finds conditions under which the participants in the society will eventually choose efficient contracts. But their actions are limited to offering contracts; no cheating occurs within a contract after it is chosen. For example, in his "marriage game," the prospective partners may both offer a contract under which they will share control, but neither is then allowed to renege and attempt to seize control. Thus issues of interest to Lawlessness and Economics do not arise. Aoki (2001) is concerned with many of the same issues of governance as I am. In his modeling of dynamics in Chapters 7–10, he sets up a general taxonomic scheme, and then interprets several examples of empirical studies in its categories. But he does not go into the details of information or communication in each game, nor does he examine how these aspects will affect the dynamics of each specific case.

Future research may achieve a synthesis that has both the dynamics and the fuller specification of the underlying games, and feedbacks between these two aspects. In

the meantime, I hope that the available approaches serve as complementary inputs to our thinking about the issues.

1.8 Approach of the Book

In the previous three sections, I have outlined the questions I will discuss in the rest of the book, and something of the methods I will use for that purpose. Here I describe and discuss my approach in greater detail.

Each of the following chapters takes up one set of issues within the range of Lawlessness and Economics. Each chapter begins with a selective overview of the literature. The existing body of research is not only multi-disciplinary but also multi-methodology. There is some statistical empirical work, and much more descriptive work consisting of case studies and ethnography. Theoretical research has lagged behind the empirical and descriptive, although the gap is closing rapidly.

This assessment may surprise some who have followed the literature on the New Institutional Economics. For example, Alston, Eggertson, and North (1996, p. 1) say that

> the field is long on theoretical analysis but short on empirical work. It is probably true that the stock of knowledge would grow faster if the new institutionalists put more emphasis on empirical work.

The difference arises, as usual, due to our different interpretations of the word "theory." A good general conceptual scheme clearly exists, and Eggertson (1996) depicts it in a diagram of boxes and arrows showing connections between institutions, organizations, contracts, and so on. However, I believe that if theory is to be useful in improving our understanding of specific institutions and helping us design or reform existing organizations and institutions, theoretical analysis must go beyond general schemata, and develop more detailed models of the specific situations and problems that concern us. Such modeling is expanding fast, but still has some catching up to do. That is where I hope to make some original contribution in this book.

Empirical research, especially case studies and ethnography, and theoretical modeling have different and complementary merits and drawbacks. Case studies give us rich, detailed description of the facts of each situation. Econometric research establishes correlations among variables across countries or over time. However, both approaches leave basic questions of cause and effect implicit or unexamined. Theoretical modeling more explicitly sets up hypotheses about causes and effects, and examines all the logical consequences of the set of hypotheses under consideration. It brings together some key aspects that are common to many situations and cases, and gives a sharper and deeper understanding of forces and mechanisms that operate. This improved understanding is what justifies theoretical modeling. I would like to steal a line from Casper Gutman in *The Maltese falcon* and claim that

formal models are "genuine coin of the realm, sir. With a dollar of this you can buy more than with ten dollars of talk" (Hammett 1930, Chapter 18).

However, for reasons of tractability, such work must abstract from many of the rich details of specific contexts. Thus each theoretical model selects for its in-depth analysis only a narrow subset of the range of facts and situations known to us from empirical and case studies. The aim of theory should be to construct a collection of models that is sufficiently small to be remembered and used, and covers a sufficiently large portion of the spectrum of facts. Overall understanding of the rich arrays of facts will best be achieved by interpreting them in the light of an appropriate model (or a few pertinent models) from this set, and by compiling and comparing the insights they generate. What about one universal model to cover all the facts? Neoclassical economic theorists who admire the generality and the beauty of the Arrow–Debreu model of competitive equilibrium may wish to attempt similar grand theorizing everywhere. But that has not proved to be a fruitful approach in other areas of economics. In the field of Industrial Organization, for example, theoretical progress has come from a toolkit of several models. I believe the same is likely to be true in the emerging field of Lawlessness and Economics. Only a small number of models covering only small patches of the territory have been constructed so far, but that simply means that there is a lot of room for further modeling.

How should one judge such models? I suggest two criteria. First, a model should do more than explain just the simplest and most obvious motivating facts that led one to build that model, it should offer a sharper or deeper understanding of them in terms of more basic economic principles or by connecting them to other seemingly unrelated facts. Second, a theoretical model should not merely reproduce as results the factual observations of case studies that the model was constructed to explain in the first place; it should yield some new results or hypotheses that can then be compared with other facts. Conversely, case study or empirical research should not treat each case as a mere narrative or description of an isolated situation; it should attempt to place it in an overall framework of other cases and theories. Ultimately progress must come from a dialogue and feedback between the different modes, not from any one of them on its own.

My comparative advantage is in theoretical modeling, so I spend more time and space on that. By way of background and motivation for the theory to come, I begin each chapter with a selective look at the literature—concepts, descriptions and case studies of some institutions and organizations, and surveys. Then I pick up some aspects of the findings of the case studies, and construct one or more theoretical models of them. I attempt to live up to the self-imposed criterion of asking the models to deliver more than just the facts they were rigged to explain; sometimes I even succeed.

In the concluding section of each chapter I assess where the modeling of that chapter gets us, what it leaves out, and what the prospects are for future research. In each chapter the readers will find several interesting empirical observations that my models do not touch at all. I do not apologize for that. This book is at best an interim stock-taking of a field in its infancy. The omitted dimensions need other models. Readers should take the frequent and large gaps and flaws I point out as challenges for further research, and an invitation to join the band of lemmings who form the field of Lawlessness and Economics. Very fittingly, no property rights to ideas are staked out or enforced in this field; everyone is free to take any idea and to run with it.