

1

What Are Institutions? How Should We Approach Them?

[I]n the great chessboard of human society, every single piece has a principle of motion of its own, altogether different from that which the legislature might chuse to impress upon it. If those two principles coincide and act in the same direction, the game of human society will go on easily and harmoniously, and is very likely to be happy and successful. If they are opposite or different, the game will go on miserably, and the society must be at all times in the highest degree of disorder.

—Adam Smith, *The Theory of Moral Sentiments* (1759:234)

Economists have traditionally been engaged in analyzing the workings and implications of the market mechanism. Markets can undoubtedly be considered one of the most salient institutions that human beings have ever produced. However, recently it has been increasingly recognized that “institutions matter” for understanding the diverse economic performances of different economies, and when the phrase is cited, the reference is not always limited to markets. Indeed, in the last decade of the twentieth century we have witnessed various institution-relevant events and phenomena that have had, and in many cases will continue to have, significant impacts on the performances of the relevant economies. There were, for example, the demise of the communist states and the subsequent transformation of their economic systems, the emergence of the Silicon Valley phenomenon and e-commerce, the European currency unification and market integration, the Japanese and East Asian financial crises subsequent to their “miracle” phases, the persistent ethnic divides and the stagnation of African economies, the global integration of financial markets and recurrent currency crises, the re-examination of the role of international organizations with nation states as members, and the growing global, nongovernmental organizations. On the surface some of these may be thought of as purely market phenomena or matters of organizational design. However, if we try to understand the causes and implications of any of these events and phenomena at a deeper level, we are compelled to take their institutional aspects into consideration.

What are institutions? Can we identify them with statutory laws, informal norms, established organizations, contracts, people’s mind-sets, or possibly combinations of some or all of these? A proper formulation of a concept, such as that of institutions, may depend on the purpose of the analysis. For example, consider the following question: If institutions matter to economic performance, why can’t the best institutions from better-performing economies be learned and imitated by other economies? This was the major issue raised by D. North in a seminal book on institutions (North 1990). To deal with it, North conceptualized institutions as the rules of the game in a society. We were told that there are two types of game rules: formal ones (constitutional, property-rights rules, and contracts) and informal ones (norms and customs). Then, even if good formal rules are borrowed from without, tension may be created

since indigenous, informal rules are inert and difficult to change. As a result a borrowed institution may be neither enforceable nor functional.

Consequently economists are becoming interested in the issue of enforceability. When do the rules of the game become enforceable? With the advent of an enforcer? But, how can the enforcer be motivated to enforce the rules of the game? In short, how is the enforcer enforced to do a prescribed job? A way out of this infinite chain of reasoning may be to show how the rules of the game are endogenously generated, and thus become self-enforcing through the strategic interactions of the agents, including the enforcer. A reasonable way of approaching institutions from this perspective is then to conceptualize an institution as an equilibrium outcome of a game. Thus we have recently seen the publication of some important works based on views of an institution as an equilibrium of a game, although most of them derive insights from historical cases (some representative works are referred to in the next section). Can we apply the same idea to the contemporary economy that appears to exist as a complex of many different institutions? Is this merely a bundle of more or less autonomous institutions, or does it exist as an internally coherent whole, that is, as an equilibrium phenomenon of some sort?

When we view institutions (and possibly their complexes) as equilibrium phenomena, this does not imply that institutions are rigidly frozen; they do change. The demise of the communist states and the subsequent transformation of the planning systems in Central and Eastern European economies is its eloquent manifestation. Then, how can one explain theoretically the emergence of an institution and/or an institutional change? In general, game-theoretic models can have multiple solutions (equilibria) and/or generate solutions highly dependent on the specification of models. Is it then that institutional emergence/change is explained merely as the selection of one equilibrium from the many that are equally possible and/or a transition from one equilibrium to another, given the fixed structure of the game? If so, is the selection/transition essentially technology—or market—induced, and does it eventually become locked in due to technological economies of scale? Alternatively, is institutional evolution programmed by “cultural genes”? Can change be engineered by political entrepreneurs or engendered by mutant entrepreneurs? Do cataclysmic political events have stochastic impacts on the selection of a new institution? Or, is there something else involved in the process of institutional change? In particular, how does the novelty often observed in the emergence of new institutions come about?

The basic research agenda of this book may be set forth as two problems: the *synchronic problem*, whereby the goal is to understand *the complexity and diversity of overall institutional arrangements across the economies as an instance of multiple equilibria of some kind*, and the *diachronic problem*, whereby the goal is to understand

the mechanism of institutional evolution/change in a framework consistent with an equilibrium view of institutions, but allowing for the possibility of the emergence of novelty.

We will investigate the institutional diversity and the complexity of economies by looking into the nature of the interdependencies of institutions across economic, political, organizational, and social domains, as well as that of institutions linking those domains. In so doing, we will reconsider the framework of traditional economics and try to incorporate some important contributions to institutional issues in neighboring disciplines, such as sociology, political science, law, and the cognitive sciences. However, departing from the old institutional economics, we will analyze the sources and implications of institutional diversity within a unified, generic—game-theoretic—framework rather than merely compiling a rich institutional catalog or drawing an ad hoc taxonomy of institutions. Developing a unified conceptual and analytical framework and incorporating important contributions from different disciplines into it will help us gain a deeper theoretical understanding of the workings of the economic institutions.

However, we also emphasize game-theoretic analysis in the traditional sense cannot be complete by itself as a systemic study of institutions. The analysis of the interdependencies of institutions within a game-theoretic framework would indicate the possibility of *multiple*, suboptimal, Pareto-unrankable institutional arrangements. That is, institutional arrangements can be diverse across economies even if they are exposed to the same technological knowledge and are linked through the same markets. Thus we need to rely on comparative and historical information to understand why particular institutional arrangements have evolved in one economy but not in others. By this we imply that an institutional analysis must be also comparative and historical, and thus we have hope to provide the groundwork for *comparative institutional analysis* (CIA).¹

In considering the diachronic process of institutional evolution, we will take an important departure from traditional game theory. Midway through the book (chapter 9) we will abandon the assumption that the players of a game have complete knowledge of the objective structure of the game they play. Instead, they are assumed to have individual, incomplete cognitive views regarding the structure of the game they play—what we call subjective game models. When actions taken by the players of the game based on their subjective game models become mutually consistent over periods (i.e., equilibrated), then their subjective game models can be confirmed by the observed reality jointly created by their action choices and reproduced as a guide for their further action choices. We will then conceptualize an institution as a salient, common component of the players' subjective game models—that is, as shared

beliefs about the structure of the game that they actually play. When action choices derived from such models do not yield anticipated results for the players, and thus a state of general perceptual crisis is created, a search for new subjective models may be triggered and continue until new equilibrium is achieved. In effect, understanding the process of institutional change may be tantamount to understanding the ways in which the agents revise their beliefs in a coordinated manner. From this perspective we can analyze the roles of technological and other environmental changes, political programs and discourses, enactment of statutory laws, entrepreneurial experiments, cultural legacies and so forth, in the process of institutional change, but this will be done after we have dealt with the synchronic problem.

The composition of this introductory chapter is as follows: Section 1.1 provides an overview of the different conceptualizations of institutions that economists have proposed. Section 1.2 introduces our conceptualization of institutions based on an equilibrium view, subject to a precise formulation later (chapter 7). Section 1.3 introduces some basic notions such as the “game form” and the types of “domains” that will play important roles in this book and then presents the plan of the book.

1.1 Three Views of Institutions in Game-Theoretic Perspective

The statement that “institutions matter” does not make much sense unless we have a common understanding about what institutions are and how they are formed. Emile Durkheim, a pioneer of modern sociology, once defined the discipline of sociology as the “science of institutions” and that of economics as the “science of markets.”² Leaving aside the old school of institutional economists, main stream economists of the past were indeed engrossed with market analysis.³ Today we see that not only economics can make significant contributions to understanding the nature, origin, roles, and implications of institutions, but important economic phenomena and problems cannot be well understood without an analysis of nonmarket institutions. Recently an increasing number of economists have taken up the task of conceptualizing and analyzing institutions. As we will see, there are at least three different (yet interrelated) meanings that economists have attached to the word “institution.” What we should be concerned with is obviously not a semantic clarification of the word as such, but a conceptualization that may be conducive to a better understanding of the workings of diverse economic systems.

In order to clarify the differences among the three meanings, or conceptualizations, of institutions that economists use, an analogy of the economic process with a game is apt. I have already indicated that the application of game theory is an indispens-

able component of comparative institutional analysis. The game-analytic apparatuses we will apply in this book to deal with the synchronic problem, namely those borrowed from the theory of evolutionary and repeated games, are relatively recent. However, the analogy of the economic process with a game can be dated back as far as Adam Smith, as the quotation in the beginning of this chapter shows. There the game is identified with a situation in which individual agents *strategically* interact with each other according to their own motivations, and this precisely corresponds to the situation with which modern game theorists are concerned. In the analogy of the economic process with a game, economists have regarded an institution as comparable to either players of a game, the rules of a game, or equilibrium strategies of the players in a game.

When people casually talk about institutions in daily conversation, they usually mean certain prominent organizational establishments. Some economists follow this convention, effectively identifying institutions as specific players of the game, such as “industry associations, technical societies, universities, courts, government agencies, legislatures, etc.” (Nelson 1994:57). But there is a second view, as North argues, that institutions should be identified with the rules of the game as distinct from its players.⁴ He opens his seminal book on institutions and institutional change with the following passage:

Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction. . . . In the jargon of the economist, institutions define and limit the set of choices of individuals. (North 1990:3–4)

Humanly devised constraints may be informal (e.g., social norms, conventions, and moral codes) or formal (i.e., consciously designed or articulated). Formal rules include political rules (constitutions, regulations), economic rules, and contracts. Economic rules define property rights, that is, the bundle of rights to use and dispose of an economic resource and to derive utility (income) from it. Contracts are (enforceable) agreements, embedded in property rights rules, regarding the use or exchange of goods. The formal rules of the economic game cannot be constructed (changed) by the players of the game themselves while they are playing. These rules are determined prior to playing the game. Since we are concerned with the origin of institutions, an immediate question arises: Who determines the economic rules? It is here that North draws a sharp distinction between the rules of the game and the players of the game (organizations and their political entrepreneurs) who can act as agents of institutional change, that is, as rule-makers. According to North, the existing rules of the game shape the incentives of the players as to how to transact and what to innovate, ultimately generating effective demands for new rules in response to changing relative

prices. The new rules are then negotiated and determined in the “political market,” that is, structured according to political rules. North claims, “[i]t is the polity that defines and enforces the property rights” (1995:23).⁵

A more technical formulation of the rules-of-the-game view is presented by Hurwicz (1993, 1996) who focuses on the issue of enforcement. In this approach the rules of a game are expressed by specifying who play the game, what actions players can choose (“a choice set”) and what physical outcome corresponds to each profile of the players’ choices (“an outcome function”). He calls such a triplet of specifications a “mechanism” or a “game form.”⁶ To illustrate, let us take the mechanism of price control whereby a seller is constrained by a ceiling set by the government on the sales price that can be charged. The constraint on his/her choice set is represented by a specific parameter value, which is, the ceiling price.⁷ According to Hurwicz, other restrictions are needed as well for arriving at a proper definition of institutions. He considers that the rules need to be enforceable, or “implementable” in his terminology. Namely he requires that only a class of enforceable, human-made restrictions on actions qualifies as an institution. He formalizes the notion of enforceability in terms of *Nash equilibrium*. A profile of strategic choices by players is said to be a Nash equilibrium if no player has incentives to change his strategy when other players are expected to remain with the prescribed strategies. In order for a set of humanly devised restrictions on the game form to be enforceable, it must then contain a Nash equilibrium as players choose strategies freely from the sets of all technologically feasible actions.

Hurwicz’s main concern is to inquire into the possibility of “designing” an institution that can implement a given social goal in a way that is compatible with the incentives of the players for a certain class of environment (technology, preferences, and resource endowments). A social goal (efficiency, equity, clean air and water, etc.) may be expressed in terms of a certain set of outcomes (consequences) to be attained for each economic environment. Suppose that a legislator designs a mechanism that implements the prescribed social goal. However, there is no guarantee that this mechanism is enforceable. For example, the legislator may expect that a price control can achieve the social goals of price stability and distributive equity, but there will always be sellers who find it appealing to sell in the black market at a price higher than the regulated ceiling price. Then price control is not self-enforceable, and thus not implementable.

If a mechanism that was designed with the purpose of achieving a prescribed social goal is not self-enforceable, then it needs to be supplemented by an enforcement mechanism. The game form must be altered by adding enforcers (the court, police, ombudsmen, etc.) with particular action sets (putting people in jail, etc.) and modifying the outcome function accordingly. But this creates a dilemma for the mechanism

designer. To make the enforcement mechanism effective, appropriate incentives may need to be provided for the enforcers to perform their mission properly. Further the operation of the enforcement mechanism may require the use of resources that have to be diverted away from activities directly contributing to the prescribed social goal. As a result the achievement of the original social goal will need to be compromised.

In considering the incentives of enforcers, Hurwicz's idea of an institution actually comes close to the third, game equilibrium notion of institutions. One of earliest proponents of this third view is Schotter (1981).⁸ More recently there have been two major developments in the game equilibrium view of an institution based on different equilibrium notions in the evolutionary game approach and the repeated game approach. Representative works of the former approach are by Sugden (1986, 1989), Aoki (1995), P. Young (1998), Okazaki and Okuno-Fujiwara (1998), and Bowles (2000).⁹ In the evolutionary game approach, a convention of behavior establishes itself without third-party enforcement or conscious design. As a convention evolves, agents tend to develop particular traits (perceptions of the environment, preferences, skills, etc.) under the pressure of evolutionary selection. Thus a convention and associated individual traits may co-evolve. A convention may eventually be codified through the judicial process to reduce the costs of disequilibrium caused by mutation and mistakes. Also an articulation in words of conventionalized rules of conduct may help make clear a particular situation. However, Sugden argues, following the tradition of Hume, that it may be misleading to think of the law as a creation of the government imposed on its citizens. Rather, "the law may reflect codes of behavior that most individuals impose on themselves" (Sugden 1986:5).

An alternative game-theoretic approach to institutions is that developed by Greif (1989, 1994, 1997b, 1998b), Milgrom, North, and Weingast (1990), Greif, Milgrom, and Weingast (1994), and Calvert (1995), among others, who rely on sophisticated concepts of equilibrium, such as subgame perfect equilibrium, in repeated prisoner's dilemma games. The precise conceptualization of subgame perfect equilibrium will be given later in this book (chapter 7). However, it may be worth noting at this point that this and other related equilibrium concepts are useful in clarifying the role of expectations or beliefs shared by players of the game. A subgame perfect equilibrium prescribes a strategy for each player constituted as a comprehensive plan of action choices contingent on all possible future states of the game.¹⁰ Any element of the comprehensive plan, that is, an action choice prescribed for a particular contingency, needs to be a Nash equilibrium when that contingency actually arises, and thus self-enforcing. As a result of applying subgame perfect strategies, some states may never be observed in the actual playing of the game. This is not because a path of play leading to such a state is excluded by exogenous constraints but because the strategic

calculations of the players mutually deter them from choosing that path once the equilibrium “plan” are put into use. Since the portions of the equilibrium strategies that prescribe actions to be taken off the paths of play are not actually observed, they may be interpreted as representing the rational expectations or beliefs held by other players regarding what actions would be chosen by the relevant players once such paths are selected in the game.

The point may be illustrated by using the model of a merchant guild provided by Greif, Milgrom, and Weingast (1994). This game is played repeatedly between a group of merchants and the ruler of a city or trading center in the context of medieval trade. In order to expand trading opportunities, the trading center needs to be organized in ways that secure the person and property of the visiting merchants. The ruler of the city might pledge that visiting merchants would be provided with this security, but once trade is established, the ruler might be tempted to renege on the pledge. Suppose now that the merchants, who have organized themselves into a guild, adopt the following strategy: they trade in the city in a given period if and only if none of them has been cheated by its ruler. Otherwise, they organize a boycott (we leave aside the matter of the guild’s ability to enforce compliance among its members). The ruler adopts the following strategy: he does not cheat unless a boycott is announced by the guild. Once a boycott is announced, the ruler cheats any trader who offers to trade. The authors proved that such a strategy profile constitute a (perfect) equilibrium. In the actual play of the game, cheating and boycotts may not be observed in normal circumstances. But this is not because they are *a priori* precluded by the rules of the game, but because the ruler expects the guild to credibly boycott him if he cheats, so it does not benefit him to do so. The formation of the guild thus functions to force the ruler to credibly commit to his pledge and thereby to allow the city’s trade expansion to proceed. In this example, the merchant guild (an organization) and its expected role of organizing a boycott in the event of cheating (the off-the-path-of-play portion of the equilibrium strategy) may be considered to provide a credible contract enforcement mechanism.

Based on this and other important works (1994, 1997b), Greif gives the following summary notion of an institution from an equilibrium perspective. Observe what this reveals the importance of beliefs and self-enforceability.

Given the technologically determined rules of the game, institutions—the non-technological constraints on human interactions—are composed of two interrelated elements: cultural beliefs (how individuals expect others to act in various contingencies) and organizations (the endogenous human constructs that alter the rules of the game [relevant to the decision-makers]) and, whenever applicable, [they] have to be an equilibrium [and thus self-enforcing]. (Greif 1994:943)

Organizations, which are social entities such as the guild in the game above, are players of a game and subject to constraints implied by an established equilibrium of the game. Greif's conceptualization may be thus said to subsume the first, player-of-the-game view as well.

Regarding the origin of an institution, we have seen that the rule-of-the-game theorists tend to subscribe to the design view; namely rule-making is susceptible to conscious design by legislators, political entrepreneurs, or mechanism design economists. Among the equilibrium-of-the-game theorists, in the beginning there was no clear consensus on this issue. Those who took the evolutionary game approach clearly subscribed to the view of an institution as a "spontaneous order" (Menger 1883; Hayek 1973) or a self-organizing system. In contrast, the concept of subgame perfect equilibrium presumes that individual players are perfectly capable of deductive reasoning regarding a possible feedback mechanism between their own and others' choices. How is it that individuals will jointly select strategies that are mutually consistent and lead to the construction of an institution, especially where there are multiple equilibria possible? There is nothing that the notion of subgame perfect equilibrium can reveal about why a certain institution evolves in one place and another evolves elsewhere. Take the example above of medieval trade where the combination of no trade and cheating in each period (and thus the observation of only no trade) can be another subgame perfect equilibrium. It seems natural then to consider that even those who adopt the superrationality notions of equilibrium, such as subgame perfection, are doing so merely to show that a certain profile of strategies (actual plays and expectations) can become self-enforceable and sustainable, *once established*.

However, there remains one paradox that has to be resolved before we subscribe to the equilibrium view of institutions. If the role of an institution is understood as being to constrain the choices of the players in one or another way, how is such a constraint found and perceived as relevant by the players? By the emergence of an equilibrium? But, then, how does each individual player find and choose an appropriate equilibrium strategy of his own before knowing the equilibrium and thus without yet being constrained by it? In other words, how can consistency be induced in the players' beliefs regarding the emergent situation and in the actual situation created by the choices of the players based on these beliefs? This question may appear to be merely about the ordinary stability property of an equilibrium. However, we will see later in this book (chapter 7) that the problem is more fundamental and cannot be resolved so simply. This is why we propose a new definition of institutions essentially based on an equilibrium view but with a substantive qualification, as introduced in the next section.

1.2 Aspects of Institutions: Shared Beliefs, Summary Representations of Equilibrium, and Endogenous Rules of the Game

Our Conceptualization of Institutions

As already noted, which definition of an institution to adopt is not an issue of right or wrong; it depends on the purpose of the analysis. Since the main objective of this book is to understand the diversity of institutional arrangements as well as the nature of the process of institutional change, we now introduce a definition of institutions that will be useful and amenable to the analysis of these issues. Because it is hard to provide a brief definition and its full implications at the outset, we will tentatively characterize the institution as a *self-sustaining system of shared beliefs* about a salient way in which the game is repeatedly played. We can identify “a way by which the game is repeatedly played” as the rules of the game. However, by that we do not mean the rules exogenously given or conditioned by the polity, culture, or a meta-game, as the rules-of-the-game theorists do. We regard these rules as being endogenously created through the strategic interactions of agents, held in the minds of agents, and thus self-sustaining—as the equilibrium-of-the-game theorists do. In order for beliefs to be shared by the agents in a self-sustaining manner and regarded by them as relevant to the consequences of their choices, they must have substantive bases. The content of the shared beliefs is a *summary representation (compressed information) of an equilibrium* of the game (out of the many that are theoretically possible). That is to say, a salient feature of an equilibrium may be tacitly recognized by the agents, or have corresponding symbolic representations outside the minds of agents and coordinate their beliefs.

By focusing on equilibrium “beliefs,” we closely follow the conceptualization of an institution by Greif, as quoted above.¹¹ However, we keep the equilibrium notion behind our definition unspecified at this stage except that it is a Nash—that is, self-enforcing. Later, in chapter 7, we will propose a more precise, encompassing definition of institutions, inclusive of both the classical and evolutionary game approaches,¹² as well as some liberal modifications. Here we specifically refer to the “summary” or “information compression” nature of institutions. As is made clear subsequently, this specification will be useful for understanding the dynamic process of institutional change. In any case, from the proposed perspective, we are concerned throughout the book with what type of institutions can become viable under what conditions and how they relate to each other. Although it is yet premature to give a precise, generic formalization of this view, we will present here a basic underlying idea to motivate our study.

To begin, let us consider a game played by a fixed set of agents, each endowed with a set of technologically (and perceptually) feasible actions. For each combination of action choices by all the agents—referred to as an action profile—a specific payoff distribution among the agents is associated. Tentatively let us refer to the collection of the sets of feasible actions over the agents as the domain of the game and the rule (function) that specifies a payoff distribution for each action profile from the domain as an exogenous rule of the game. Given these characteristics of the game, each agent wants to maximize his/her payoff. However, his or her payoff cannot be solely determined by his/her action. The best action choice of each agent ought to be contingent on others' action choices, but how can agents infer others' action choices?

Suppose, for a moment, that a game is played repeatedly, out of which a stable outcome (an action profile) somehow evolves and each agent has come to have a reasonably good idea, based on his/her experience, about how the game is played in the domain. Namely agents may not be able to infer, or may not even need to infer, every detailed characteristic of the others' action-choice rules but come to perceive some salient features of private rules that relevant agents are believed to apply in making their action choices. Relying on such compressed information, each agent may also develop his/her own private rules—strategies—for making an action choice in response to an evolving state of the domain. Clearly, a complex feedback mechanism is operating here. All the agents form their own action-choice rules as their strategies in response to their subjective perceptions (beliefs) of others' action-choice rules even though in an incomplete and compressed form. Only when their compressed perceptions about others' action-choice rules become stabilized and reproduced, can their own action-choice rules also become stabilized and serve as useful guides for playing the game, and vice versa.

We may capture this consistency property evolving with respect to agents' beliefs and their strategic formation of action-choice rules by regarding them as being in (Nash) equilibrium. It is not beneficial for the agent to deviate from his/her own action-choice rule thus constructed, as long as their beliefs regarding other agents' action-choice rule are sustained. Institutions can then refer to that portion of agents' equilibrium beliefs common to (almost) all of them regarding how the game is actually played (how action-choice rules are applied by agents in the domain). Although they are equilibrium phenomena, they should be regarded neither as a result of perfect deductive reasoning in a one-shot game, nor a complete stasis to which no inductive reasoning needs to be applied by agents. They represent the substantive, self-sustaining expectations of the agents who have actually played the game repeatedly. As such, an institution is “the product of long term experiences of a society of boundedly rational and retrospective individuals” (Kreps 1990:183).

An equilibrium state is a socially constructed reality, and thus it is endogenous to the domain. It coordinates the beliefs of agents through its summary representations —tacit and symbolic. As an equilibrium phenomena, an institution cannot be ignored by any agent as far as others do not ignore and thus influence their strategic choices. Agents' strategic choices made on the basis of shared beliefs jointly reproduce the equilibrium state, which in turn reconfirms its summary representation. Thus the institution becomes self-sustaining and information compressed in it becomes taken for granted by the agents unless some events shaking the shared beliefs occur (see figure 1.1, where the dashed-line box represents an institution; we will momentarily ignore the dotted-line box). In this way, although endogenously created, an institution becomes objectified.¹³ By relying on equilibrium analysis, we can understand this dual nature of institutions, endogenicity and objectivity, which may have been responsible for the somewhat confused bifurcation of endogenous versus exogenous rules-of-the-game views of institutions.¹⁴

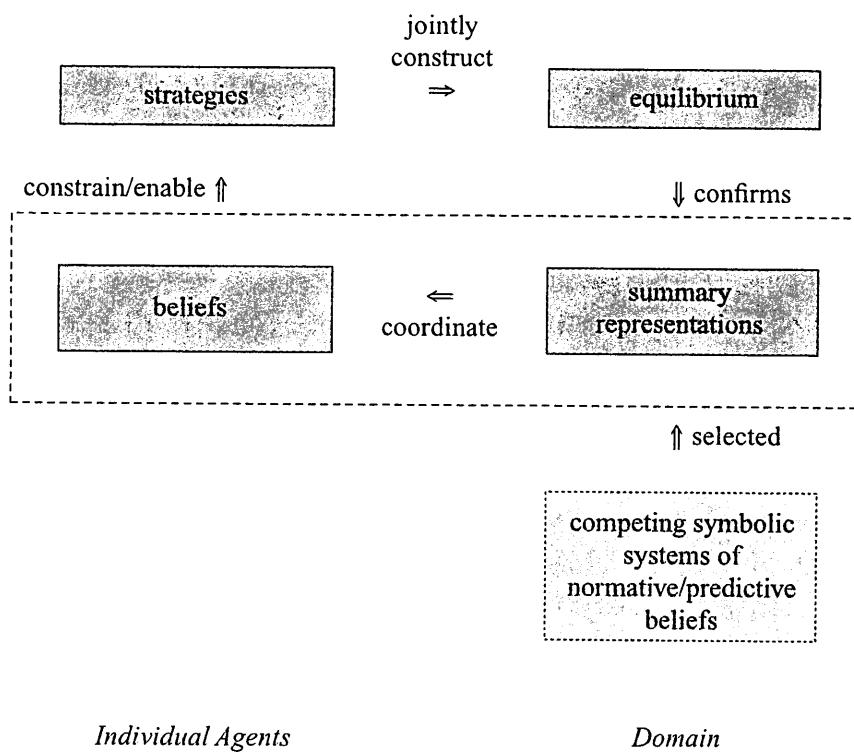


Figure 1.1

An institution as shared beliefs formed as summary representations of an equilibrium. An institution is represented by the broken-line box.

A critical feature of an equilibrium recognized as an institution may sometimes be represented in some explicit, codified and/or symbolic form, including statutory laws, agreements, social structures or organizations as systemic arrangements of differentiated roles, and so on. However, the point is that such a representation is an institution only if the agents mutually believe in it.¹⁵ From this perspective, statutory law and regulations *per se* are not institutions if they are not necessarily observed. For example, even if the government prohibits the importation of some goods by a statutory law, but if people believe it effective to bribe customs officers to circumvent the law and make it a prevailing practice, then it seems appropriate to regard the practice rather than the ineffectual statutory law as an institution. On the other hand, certain practices, if not formalized, can be institutions as long as the agents believe in them as relevant representations of the internal state of the domain; they cease to be institutions when the agents' beliefs in them are critically shaken.

The equilibrium–summary-representation view of institutions helps clarify their dualistic constraining/enabling nature. The role of institutions is normally understood as a nontechnological constraint on the action choices of the agents by the exogenous rules-of-the-game theorists as well as the equilibrium-of-the-game theorists (recall the definitions by North and Greif discussed above). Indeed, an institution, by the very fact of its existence, controls agents' individual action-choice rules by coordinating their beliefs. These beliefs channel their actions in one direction against the many other directions that are theoretically possible (i.e., other equilibria). In this sense, controlling or constraining character is certainly inherent in institutionalization. However, an institution coordinates agents' beliefs only in summary and shared ways. In a world of incomplete and asymmetric information, an institution "enables" the bounded-rational agents to economize on the information processing needed for decision-making (see figure 1.1).¹⁶

Here, an analogy with the price mechanism familiar to economists may be useful. In the market mechanism, individuals do not need to know every detail of the internal state and external environments in which they make their choices, but only the relative prices (Hayek 1945). Leaving aside the problem of the enforcement of contracts and property rights, if there were a complete set of markets, relative prices could be formally regarded as "sufficient statistics" summarizing the data (preferences and technological possibilities of production) needed for the society to achieve the social optimum in the most efficient way. The dimensionality of relative prices does not exceed the number of goods traded minus one with one particular good serving as a numéraire (Koopmans 1957; Hurwicz 1960, 1973). Of course, in actuality markets are far from complete. Individual agents need alternative means to gain

the useful information for making their choices. Various institutions other than markets then evolve in response to the failure of complete markets to exist (Arrow 1998). Thus individual agents are not only constrained but also informed by institutions. Just as markets transmit information regarding the economic environment (technologies, tastes and resource endowments) in the summary form of relative prices representing the marginal rates of substitution/transformation so do other institutions in alternative summary forms (chapter 6.2).¹⁷ Also a summary representation of an equilibrium can be robust to the mildly changing environments of the domain, as well as the associated shift in equilibrium, because of its very nature of information compression.¹⁸ Thus information compression embodied in an institution will make it possible for boundedly rational agents to efficiently collect and utilize the information necessary for their actions to be consistent with changing internal and external environments.

Five Reasons Why the Proposed Conceptualization Is Amenable to Our Analytical Purposes

A conceptualization of institutions is of course a matter of the theorist's taste and not a matter of right or wrong. However, in my view, there are at least the following five reasons why the shared-beliefs cum equilibrium–summary-representation view of institutions is useful for comparative institutional analysis. The first three points are applied to the institution-as-equilibrium-of-the-game view in general, while the last two, more or less specifically, to the institution-as-equilibrium–summary-representation view that emphasizes the cognitive aspect of institutions.¹⁹

Endogenous Treatment of Origins of Institutions and Enforcement The institution-as-an-equilibrium approach in general can deal with the issues of the origins of an institution and its enforcement endogenously. As we have seen, if one subscribes to the exogenous rules-of-the-game view, then one must immediately face the issues of where and how the rules originated, as well as how they are enforced. An institutional origin may need to be found outside the domain of the economy in which the rules are applied: for example, in the polity domain or, theoretically, in the domain of a metagame in which rational agents collectively choose a rule from the set of many possible rules.²⁰ But how, then, are the rules of the game in the polity domain set? How are all the possible rules known to the players of the metagame, and how do they play the metagame? Where are the rules of the metagame determined? Thus a problem of infinite regression is bound to arise. Perhaps the right way to partially resolve this problem is to regard an institution as originating as a stable endogenous product of the game—in economic, social, or political exchange domains—while

leaving the nontechnological rules of the game unspecified as much as possible at the outset.

One caveat is due, however. Although we wish to understand an institution arising in one domain as the “endogenous rules of a game” generated in that domain, we cannot build a model to make every possible institution simultaneously endogenous. In other words, what constitutes the “exogenous rules of the game”—namely the set of agents, the set of their action choices, the way in which each profile of agents’ action choices is transformed into consequences—may not be completely described by technology, resource endowments, and the preferences of the agents alone. This is the point first addressed by Field (1979, 1981). To see the same point from a slightly different perspective, imagine hypothetically that the exogenous rules of some game are completely specified by technology. Even if it is possible to do so, however, there will be multiple equilibria in a repeated game situation, and as already pointed out, which equilibrium is chosen from among the many cannot be determined endogenously. We need to consult historical events and rules as well as rules prevailing in surrounding domains.²¹ It may be that the particular subsets of actions that agents perceive as the sets of viable options are constrained by historical precedents, while the way in which the consequences of a certain profile of agents’ actions are determined in one domain are affected by the institutional environments of the domain (i.e., endogenous rules of the game prevailing in surrounding domains). One can never have an institution-free world from which to start the analysis and completely eliminate appeals to exogenously given, humanly devised rule structures. Thus nobody can escape from the problem of infinite regression. However, we may seek to direct the infinite regression toward structures inherited from the historical past rather than the logical construct of the metagame.

A similar problem of infinite regression can arise with respect to enforcement in the exogenous rules-of-the-game approach. Leaving aside norms and conventions that are self-enforcing (“informal” rules in North’s sense), if the rules of the game (“formal” rules) are to be enforced by an augmented player (enforcer), the question of the enforcer’s motivation needs to be addressed. Who enforces the enforcer? That is, do we need yet another enforcer to monitor the rules of action prescribed for the original enforcer? As the preceding discussion regarding Hurwicz’ contribution suggests, a solution to this problem is again to analyze a game that includes the enforcer as a player, and to see if the prescribed rules of action for the enforcer can become his/her equilibrium strategic choice and thus self-enforcing, given an equilibrium constellation of strategic choices by other agents, and vice versa. In this case, too, the presence of the enforcer as a player of the game at the outset is presumably given by history.

History Matters Through the occurrence of multiple equilibria in specific models, the institution-as-an-equilibrium approach can shed light on the “humanly devised” (North 1990) nature of institutions rather than its ecologically, technologically, or culturally driven aspects. If there is only one equilibrium corresponding to the technological specification of the structure of the game, then that equilibrium is little more than a representation of the technological condition, and not an institution.²² For example, often the evolution of community norms in East Asia is attributed to the climatic and ecological conditions there, which presumably make peasant family farming and collective use of the irrigation system more productive. However, Korea and Japan, which are characterized by similar ecological conditions, had subtly divergent institutional evolutionary paths in terms of village social structure and social norms. These sociological factors have had profound and long-lasting impacts on the subsequent institutional trajectories of both economies (chapters 2.2).

Usually, a multiplicity of equilibria is regarded as troublesome by game theorists, and they have spent many research efforts, without decisive success, in the so-called “refinement” of the equilibrium to enable them to identify only one equilibrium out of the many possible Nash equilibria. However, we consider that the multiplicity of equilibria of games should not be regarded as bothersome in comparative institutional analysis. On the one hand, by making institutions susceptible to equilibrium analysis, it can be made clear that institutions are humanly devised but can be neither arbitrarily designed nor discretionary implemented. On the other hand, once an institutional bifurcation occurs, even if two economies are exposed to the same technological and market environments afterwards, the subsequent overall institutional arrangements of the two economies may well differ, depending on their respective interim institutional trajectories. This phenomenon is known as the path dependence (David 1985). Thus equilibrium and historical analyses are mutually complementary and are both indispensable to comparative institutional analysis.

Given the impossibility of identifying every institutional phenomenon as an endogenous outcome at the same time, Greif (1998b) proposes the following an analytical procedure for dealing with historical information in the equilibrium-based approach to institutions: First, using historical and comparative information, sort out what technological and institutional factors can be treated as “exogenous” and what institutional factors are to be treated as “endogenous,” that is, must be explained. Then, build a context-specific, game-theoretic model in which those exogenous factors define the exogenous rules of the game and solve for possible equilibria. Next, find out if some of these solutions are useful for understanding the nature of the institutional factors needing to be explained. Finally, examine what “historical”

factors can be considered responsible for the selection of that particular equilibrium solution to determine the role of history.

Interlinkages and Interdependency of Institutions The institution-as-an-equilibrium approach provides an analytically tractable conceptualization of the interdependencies of institutions operating within the economy. When the government drafts a statutory law for the purpose of introducing an “institution,” its implementation may have unintended consequences in particular economic, political and social contexts. Take the example of a postcommunist economy where the government drafts a privatization law aimed at emulating markets for corporate control in an advanced economy. An outcome may be the widespread capture of corporate control by insiders, such as ex-industrial bureaucrats, directors of ex-state-owned enterprises, who amassed *de facto* control rights before the transition to a market economy.²³ This situation is somewhat analogous to the one in which a medicine tested in a laboratory has unpredicted side effects after it is administered to a human being because of the complexity of living organic systems. A major reason for such unintended outcomes is the absence of “fits” between the designed plan and the existing institutional environments that reflects a unique historical trajectory of institutional development. This suggests the possibility that only institutional arrangements that are mutually consistent and/or reinforcing may be viable and sustainable in an economy. Otherwise, an attempted institutional design may be highly unstable. It may not be accidental that co-determination in the corporate governance domain and social democratic corporatism in the polity domain co-evolved in Germany, while the main bank system, the lifetime employment system, and the close alliance between industrial associations and relevant administrative bureaus co-evolved in Japan, both in contrast to the so-called Anglo-American model (chapters 11 and 13).

We will consider institutional interdependencies as institutionalized linkages and institutional complementarities in part II. These intuitively appealing concepts are amenable to rigorous analysis when the equilibrium-oriented notion of institutions is applied. Specifically, we look at games in different domains of the economy, including organizational coordination, commodity trade, transactions of services of human and financial assets, political-transactions, and social-exchange. Then, in applying an analytical technique developed by Topkis (1978) and Milgrom and Roberts (1990), we analyze how an equilibrium constellation of strategic choices of agents in one domain can become strategically complementary to, or conditional on, the equilibrium choices of other agents in the same or other domain. In this way we can come to understand the conditional robustness of an overall institutional arrangement of the economy as well as the multiplicity of such arrangements.²⁴

From this systemic point of view, both the usefulness and the limit of agency theory as a tool for comparative institutional analysis may be touched on. Agency theory casts the economic interaction of agents (in the generic sense) in a certain domain of the economy as a principal–agent relationship. Then it inquires into what type of self-enforceable (incentive compatible) arrangement can be established as a second-best response to environmental and incentive constraints when information asymmetry exists between the principal and the agent. However, the solution is usually responsive not only to the technological environment but also to the “institutional environments” hidden in parameters specifying the objective functions of the principal and agent, and the participation constraints describing the outside options of the agent. Thus caution should be taken in utilizing and interpreting the results of principal–agent models. These results may be valid only relative to an implicitly assumed institutional environment of the domain, and may not be exclusively technology-determined, second-best solutions applicable anywhere. A rough analogy may be drawn with the relationship between “partial” equilibrium analysis of individual choice behavior with prices as exogenous parameters, and “general” equilibrium analysis of market price determination in Walrasian economics. Agency theory provides a powerful partial equilibrium analysis of an institution in a particular domain of interaction between the principal and the agent (s), with institutional arrangements in other domains taken as given environments.²⁵ However, in order to really understand why a particular institution emerges in a domain of one economy but not in a similar domain of another economy, we need to make explicit the mechanism of interdependencies among institutions across domains in each economy.

Institutional Change through the Competition of Symbolic Systems of Beliefs The equilibrium–summary-representation view of an institution suggests a new way to approach the mechanisms of institutional change. As was mentioned earlier, the information transmitted by an institution is never complete. But for the bounded-rational individual agents the compressed information may be adequate to make mutually viable choices under normal circumstances. They can still be guided by it in developing whatever skills or dispositions are in keeping with the endogenous rules of the game. However, when the pattern of choices becomes problematic because of environmental and internal changes, an “institutional crisis” in the cognitive sense may be triggered: the shared beliefs regarding the ways in which a game is played may begin to be questioned and the agents may be driven to reexamine their own choice rules based on new information not embodied in existing institutions.

A new institution will emerge only when agents’ action-choice rules become mutually consistent in a new way and their summary representation induces convergent

beliefs among them. But such a transition may not be just a move from one equilibrium to another for a given structure of the game. Rather, it may involve a novelty that cannot be characterized simply by a move from an equilibrium under given sets of action-choice rules of agents to another equilibrium under the same sets of action-choice rules (chapter 9). In the transition process, various choice rules involving new actions may be experimented with and put into competition by agents. How can the convergence of beliefs and the coordination of new choices be simultaneously induced in such a situation? As we will see later (chapter 7), the present state of economics has not been able to show that dual convergence, both in actual choices and beliefs, is possible through a reasonable mechanism of mutual interactions (i.e., actual choices are induced by beliefs and beliefs are formed by observations of actual choices), particularly when any novelty in action choices is involved.

But it can be through the guidance of a particular symbolic system of predictive/normative beliefs among the many competing ones presented in the transition process and recognized as “prominent” or “salient” that agents’ new strategic action-choice rules are forced to coordinate (Schelling 1960). As agents’ choices equilibrate, a guiding symbolic system becomes consistent with, and reconfirmed by, their experiences. It then serves as their summary representation of equilibrium incorporated into agents’ stable beliefs, namely as an institution (as indicated by the line from the dotted box to the dashed box in figure 1.1). The point is that some symbolic system of predictive/normative beliefs precedes the evolution of a new equilibrium and then becomes accepted by all the agents in the relevant domain through their experiences. It could be “unsettled culture or ideologies—explicit, articulated, highly organized meaning systems—[that may] establish new styles or strategies of actions” (Swidler 1986:278), “an entrepreneur’s vision that may trigger certain actions that eventually remove the limits of organizational capabilities and environmental constraints” (Fujimoto 1999:10), or even the political program of a subversive political party (e.g., “all factories to the workers! all the lands to poor peasants!”). In chapter 9 we will describe how bounded-rational, individual agents form their own “subjective models of the game” that they play, and discuss the mechanism of institutional change as a process of revision, refinement, and inducement of mutual consistency of such models incorporating a (common) representation system.

The Role of Statutory Laws and Public Policy Discourses Whether the rules of the game constituting institutions are viewed as endogenous to the domain or whether they are exogenously set in the polity domain may have significant implications for interpreting the role of public policy. If one subscribes to the view that institutions are made of polity-determined rules and matter to the performance of an economy,

the implications are that a badly performing economy can reform itself by government designing and implementing better rules, possibly emulating best practices elsewhere. If this is not realized, blame could then be placed on the government. However, there are two problems with this kind of argument. First is that the government itself is an organization of the people who have their own motivations and aspirations. It is an endogenous player of the game in the polity domain and the outcome of any policy-making should be understood as determined by the interactions of the strategic expectations among the players, the government, politicians, and private agents.²⁶ Blaming this on the incompetence of politicians may not pinpoint the real problem.

Second, as already noted, a policy may not yield the outcome intended by the government or politicians if it does not fit with existing institutions in other domains, an accumulated stock of competent agents, and so on. In this book we will treat statutory laws and regulations as exogenously set parameters for defining game forms (exogenous rules of the game), and examine what the outcomes will be of the strategic interactions of the agents under them.²⁷ Statutory laws or regulations may induce an institution to evolve, but they themselves are not institutions. Also an institutional outcome may be different from what a legislature or government initially intended. A careful and systematic study is called, including an investigation of how the initial “institutional” conditions, such as the legacies of old institutions and the prevailing informal rules (norms, social ethics, etc.), kinds and level of the existing stock of human competence, and so on, affect subsequent institutional change, how rule-setting in the polity interacts with the evolution of the endogenous rules of the games in other domains, and so on.²⁸

A subtle issue is that endogenizing the government does not necessarily mean that the outcome of a game will be fully determined, leaving no scope for policy advise—a paradox referred to as the “determinacy paradox” (Bhagwati, Brecher, and Srinivasan 1984). In an “institutional crisis,” individual agents may not have clear expectations about the state of the game, or even if they think they do, their beliefs may not necessarily be mutually consistent. Then there may be latitude for exogenous symbolic systems of predictive/normative beliefs to compete for the position of an attractor or a “focal point” (Schelling 1960) for the formation of coordinated beliefs. The system could well be a program or platform of competing political parties, professional political advice, an “elites’ pact” (Weingast 1997), or the drafting and enactment of statutory law. Thus political discourses in and out of polity may have a certain imprint on subsequent institutional evolution.²⁹ No matter what the competing exogenous symbolic system may be, a crucial factor that will determine its impact is the “fits” with emergent practices in domains other than the polity.

1.3 Organization of the Book

Domains, Game Forms, and Institutions

We have introduced the basic conceptualization of institutions to be developed in this book. We plan to apply this concept and examine its implications for some important contemporary and historical institutional issues from a comparative perspective. In doing so, we will develop a unified framework for analyzing the interdependencies of institutions clustering in various economies. This analytical framework will incorporate contributions and insights from economics, as well as sociology, political science, law, and the cognitive sciences, wherever possible. Before we can develop such a framework, we first need to define the terms and concepts, such as domains and the game form, that will be used in our framework, and to specify some basic types of domains.

We will treat the domain of a game as a unit of analysis. The *domain* of the game is composed of a set of agents—either individuals or organizations—and sets of physically feasible actions open to each agent in successive periods.³⁰ A combination of actions chosen in one period by all the agents in the domain is termed an *action profile*. An action profile determines the distribution of the payoffs among the agents in the domain. We decompose the payoff functions—a rule assigning a payoff distribution for each action profile—into objective and subjective elements. Namely, given external environments and historically determined states of the domains at the beginning of a period, an action profile in that period first generates a consequence in the state space describing all possible physical states relevant to the welfare of the agents in the domain. The consequence of this period defines an initial state of the succeeding period. The function (rule) that assigns a physical consequence in the state space for each action profile and a historically given initial state is called the *consequence function*. Various environmental factors, such as technology and “institutions” prevailing in other, relevant domains, as well as statutory laws and policy determined in the polity domain, parametrically define the form of the consequence function. A domain and an associated consequence function specify a *game form*, which represents the *exogenous rules of the game*.

Each agent in the domain has a preference ordering over possible consequences for each period in the state space. The composite of the consequence function and the agent’s preference function defines the payoff function of that agent in the ordinary sense of game theory. The reason we decompose this into the objective consequence function and the subjective preference function is to specify the notion of the exogenous rules of the game, as distinguished from endogenous rules of the game. Also in

games that we will discuss in this book, we do not necessarily assume that the agents have knowledge about subjective preferences of other agents, and thus they may be guided in their action choices only by objective states (physical consequences) they can observe. If all the agents in a domain choose their private plans of action so as to maximize their current payoffs, or the present-value sum of their current and future payoffs, subject to their own expectations regarding others' strategic choices, the situation is characterized as a game and the agents may be interchangeably called the players. An action plan thus chosen by an agent is his/her *strategy*. A strategy may prescribe only one action or a comprehensive plan of actions contingent on the state of the domain (or past history of states).

We deal with six basic types of domains: commons, trade (economic exchange), organization, social exchange, polity domains, and generic organizational fields. We investigate what types of institution/convention can become viable within and across these domains. As mentioned in section 1.2, North excluded organizations from the category of institutions. This is because he is interested in understanding the role of organizations as agents of institutional change in the polity domain. However, this book will place relatively less emphasis on politically determined formal rules, or on the causality from the polity to the economic process. Admittedly the mutual feedback mechanisms between the outcomes of games in the polity domain and those in the economic-transaction domain are important, and we will place substantial emphasis on them and on organizational conventions of business enterprises and other private-order entities (e.g., financial intermediaries) as integral, endogenous elements of the overall institutional arrangements of the economy.

We distinguish the above-mentioned basic types of domains primarily according to the variability of the set of agents, and the nature of the choice sets across the agents. We try to do so only in terms of their technological properties in order to identify institutions that may endogenously evolve in each domain or across domains. However, in principle, it is impossible to start an analytical discourse on institutions in a purely institution-free, technology-only setting. Thus, in the following classification of the domains, some generic institutions, such as ownership, role-based expectations, or power distribution, are inevitably present or implicit, albeit in a primitive form.

Commons Domain The set of agents in this type of domain is composed of those using common resources accessible by any one of them (and produced jointly by them). It is assumed to be technologically costly to exclude potential beneficiaries from obtaining benefits from the common resources, so the set of agents is assumed to be fixed. They may use common resources simultaneously or sequentially but not

necessarily jointly.³¹ We do not assume that the agents have a common objective or internalize a common value regulating the use of the common resources, but that they are strategic players interested in maximizing their individual payoffs subject to the expected choices of others. Since the presence of common resources is a defining characteristic of the domain, the action sets of the agents are essentially symmetric in that they all contain actions related to the production, maintenance and/or use of the common resources (e.g., contributing to the accumulation or depletion of the common resources, or using them at various intensity levels). Because of the fixity of the set of agents, individual action choices cause external economies or diseconomies throughout the domain (e.g., congestion in the use of common resources, free-riding in maintenance efforts, or benefiting others by producing common goods), but the agents may not be able to exit from the domain or to be excluded from it. An endogenous institution that may arise in this domain could be (customary) property-rights rules and group norms (chapter 2).

Trade (Economic Exchange) Domains The domains of games of this type are composed of agents endowed with privately owned economic goods that they can trade at will. Although they are initially endowed with these goods in diverse patterns, their choice sets are qualitatively symmetric in including physically possible offers of various quantities of goods in exchange for specific quantities of other goods or money, acceptance/rejection of others' offers, and honoring/defaulting on agreed-upon terms of trade. One of the important characteristics of trade games is that all agents have an option of not trading. These domains may be more specifically differentiated into the financial transaction domain, labor transaction domain, supply domain, product market domain, and so on. Domains of this type may evolve, by itself or in conjunction with other types of domains, institutions dealing with problems arising from information asymmetries between trading agents that may otherwise lead to the breakdown of trade opportunities (chapter 3).

Organization Domains In domains of games of this type, agents are able to produce goods (e.g., revenues) by their joint actions and distribute them among themselves. Although these joint actions may involve the use of some common resources (e.g., goodwill, accumulated information, and organizational infrastructure), domains of this type can be distinguished from the commons domain in two respects. First, it is optional for agents to participate in this type of game. In other words, if a game of this type is played repeatedly, agents have an option to exit, or an agent has an option to exclude others from the domain at the end of any period, and thus the set of agents is not fixed. Second, the sets of actions may be substantively differentiated across agents on the basis of the division of operational and cognitive labor (e.g., into

managerial, engineering, and operational tasks), but there is a focal (centralized) agent—the management—whose role is to coordinate the choices of the agents in the domain.³² Trade domains and organizational domains share a common characteristic in that participation is optional for agents. However, while coordination in trade games can be achieved through a multitude of voluntary agreements between two traders, in organizational games systemic incentives need to be provided to all agents in order to induce their participation and properly coordinate their action choices (chapter 4).³³

Organizational Field We also consider a generic type of domain called the organizational field, which embeds individual organization domains. This is the relatively unstructured, primitive domain in which organizations are created by the matching of agents from the population of the domain. Depending on matched types of human assets that agents have invested for information processing, different types of organizational architectures may be generated. Agents may withdraw from this domain by choice, but they are assumed to be symmetric in their action choice sets. That is, they may choose types of human assets and decide whether or not to accept a particular matching, but no hierarchical assignment is specified prior to matching. This domain is conceived of as a theoretical construct, useful for understanding the logic involved in the co-evolution of a convention of organizational architecture and a type of human assets (chapter 5).

Polity (Political Economy) Domain The set of agents in this domain contains a unique focal (centralized) agent—the government—endowed with a set of action choices asymmetric to those of the other agents who are called private agents. The latter can be citizens, interest groups, business associations and unions, economic classes, and so forth, depending on the context. The government's set of actions may include the unilateral extraction (transfer) of properties from private agents to itself or to other agents (e.g., taxes, subsidies, or fines), the compulsory mobilization of services of private agents (e.g., military and jury services), the organized infliction of physical violence on private agents (e.g., death penalty or arrest), and the monopolized supply of public services such as law enforcement. Private agents cannot escape from government action by choice, and thus the government has exclusive regulatory power. However, private agents may choose to support or not support (resist) the government. If government action invokes strong resistance from private agents, the consequence can be costly to the government (e.g., the loss of power), although resistance may be costly to the private agents as well. When a stable outcome is observed in a polity domain in which the focal agent is identified with a national government, we refer to its salient properties as a nation *state*. In other words, we

distinguish the government as a player of the game and the state as a stable outcome of the game (chapter 6).³⁴

Social Exchange Domain This domain plays an important complementarity role in understanding institutions, such as community norms, status differentiation, rank hierarchies within homogeneous teams or organizations, and so on.³⁵ In this domain noneconomic goods/bads (social symbols, languages) that would directly affect the payoffs of recipient agents, such as esteem, approval/disapproval, sympathy, accusation, benign neglect, and so on, are unilaterally delivered and/or traded with “unspecified obligations to reciprocate” (Blau, 1964/1988), and sometimes accompanied by gift-giving.³⁶ When exchanges are multilateral and diffusive among a fixed set of agents who are mutually identifiable, we call it a community. It can be the rural community, the community of traders, a professional community, and so on. These domains generate various types of social norms in conjunction with other types of domains (chapter 2).

Figure 1.2 provides a diagram of these six types of domains. The vertical dimension refers to the qualitative variation of the action sets across the agents in the domain; the horizontal distinguishes whether the agents have the option to exit, or to be excluded, from a game. The location of the six types should be evident from this characterization.

Since any classification of domain types cannot be made in purely technological terms, the domains presented above cannot be sharply delineated. If we take a particular set of agents who are strategically interacting with each other, classification

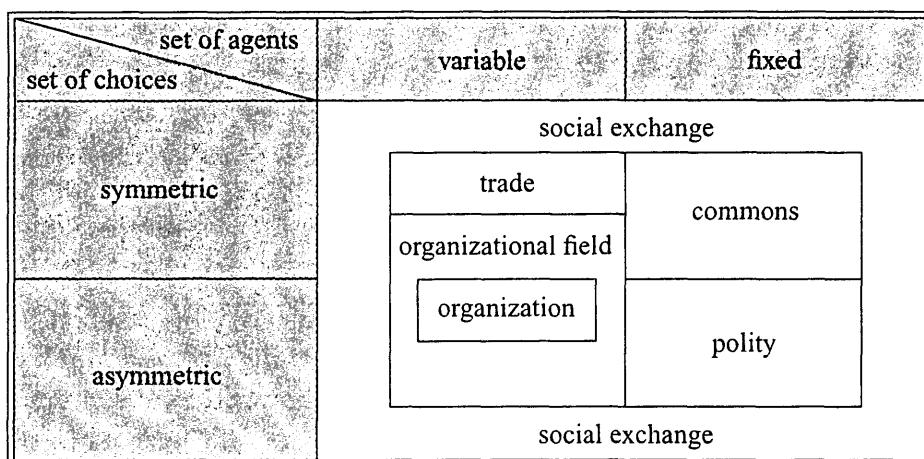


Figure 1.2
Six types of domains of games

may not neatly fall in with one of the types, but have multiple domain characteristics. Take the classification of a firm. A firm's most salient characteristics are organizational in incorporating a division of operational and cognitive labor. However, some contract-theory economists focus on the aspect of the firm as a "nexus of contracts" (Jensen and Meckling 1976): the aspect that can be understood as an institutional arrangement in the trade (economic exchange) domain. Generically, the evolution of a convention regulating the internal coordination of the firm may be captured as an outcome of an evolutionary game in the organizational field. The corporate firm also has a community aspect in which social exchanges take place among its members to form the corporate culture (e.g., "IBM man") and various suborganizational norms (e.g., shop-floor work norm). It has a modicum of the notion of a commons domain as well (e.g., in the use of intangible information assets). Finally, the firm is embedded in some governance structure that resembles an institution in the polity domain (corporate organ, workers cooperatives, kibbutz, partnerships, etc.). We cannot develop a theory of the firm taking into consideration all these characteristics at once from the beginning. In various parts of the book, we will focus on one aspect or another. However, our ultimate objective is to understand the logical structure in which the various facets of a business firm fit each other in alternative ways, depending on historical and environmental contexts.

Now let us reiterate the intuitive concept of institutions based on the shared-belief cum equilibrium–summary-representation perspective, subject to its formal and substantive refinement later in the book. Suppose that agents choose their action-choice plans strategically in a domain or across domains and that a stable outcome evolves in that domain or across those domains and is sustained over time. Then, provided that there is another equilibrium (or more generally, another sequence of equilibria), we identify an *institution* as follows:

An institution is a self-sustaining system of shared beliefs about how the game is played. Its substance is a compressed representation of the salient, invariant features of an equilibrium path, perceived by almost all the agents in the domain as relevant to their own strategic choices. As such it governs the strategic interactions of the agents in a self-enforcing manner and in turn is reproduced by their actual choices in a continually changing environment.

Five elements are present in this conceptualization: *endogeneity* (as implied by "self-sustaining," "self-enforcing," and "reproduced"), *information compression* (as implied by "a compressed representation"), *robustness* with respect to continual environmental change and minor deviance (as implied by "invariant features of an equilibrium path," "perceived by almost all the agents" and "reproduced ... in a continually changing environment"), *universality* of relevance (as implied by "shared

beliefs”, “govern the strategic interactions of the agents” and “perceived by all the agents”), and *multiplicity*. Depending on whether or not a domain is symmetric with respect to the action sets of agents, an institution can summarily represent distinct action-choice rules for different types, or identical rules for all the agents in the domain. In the polity domain, where the action sets are asymmetric across the agents, summary representations are comprised of expectations of distinct state-contingent action choices by the government and citizens/interest groups (recall the example of the merchant guild). By the same token, in the organizational domain, they can be comprised of distinct roles expected for agents occupying different positions (e.g., manager, foremen, workers) in the organization. On the other hand, in the commons and trade domains where the action sets are symmetric across the agents, they take the form of norms and self-enforcing contracts supported by certain shared beliefs about ways in which the game is repeatedly played.

The economy can be considered as constituted of myriad domains: commons, economic and social exchange, organization, and polity, some of which overlap, some of which are nested in others, and so on. For the same exogenous rules of the game (e.g., technology), multiple institutions are possible in each of these domains. Not only that, but institutions can evolve across different domains, linked by the coordinated strategic choices of agents. We refer to a synchronous set of institutions across constituent domains in the economy as an *overall institutional arrangement*. Needless to say, their structures can be very complex but not necessarily randomly figured. One purpose of this book is to discover the generic laws of regularities that prevail across different overall institutional arrangements.

The Plan of the Book

This book is composed of three parts. Part I takes up successively the six types of domains in primitive form and identifies the prototype institutions as stable multiple equilibria of games in each domain. This provides a foundation for the generic, game-theoretic framework constructed in part II for conceptualizing institutions, analyzing their interdependencies across domains and over time, and thereby understanding the mechanism of institutional change. Of course, we need to check if this framework is useful in order to understand the complexity and diversity of the institutional arrangements of the contemporary economies and their changes. This is done in part III.

Part I begins with chapter 2, which deals with the commons domain and the social exchange domain that embeds it. It derives the customary property-rights rules and a community norm as endogenous outcomes of the strategic interactions of the agents

in those domains. Chapter 3 is concerned with trade (economic exchange) domains and derives various autonomous institutions that may govern and enhance trade and markets without the intervention of the government based on the rule of law. Chapter 4 focuses on the organization domain. It identifies the various organizational and quasi-organizational architectures of practical relevance, discusses their relative information efficiency, and examines their governance issues. Chapter 5 deals with organizational fields in which different organizational architectures can co-evolve as a convention with associated types of human assets (information-processing competence) and discuss ways in which gains from organizational diversity are exploited. Chapter 6, the end of part I, turns to the polity domain and identifies various types of states as stable equilibria of the political exchange game with the government as a player. Institutions identified and discussed in part I are referred to as *proto-institutions* because they are introduced one by one, in an inevitably primitive form and without an explicit analysis of the interdependencies among them. This part largely relies on the work of many authors, including myself, and provides an overview of the current state of analytical approaches. It is by no means intended to be a comprehensive survey, however.

Building on the foregoing preparatory, taxonomic analysis, part II is devoted to the construction of a generic analytical framework for institutional analysis. Illustrative examples and cases help provide the basic motive behind this unified framework within which the systemic nature of overall institutional arrangements of the economy, as well as their changes, can be analyzed. Chapter 7 takes a precise game-theoretic conceptualization of institutions as a self-sustaining system of shared beliefs cum summary representation of equilibria, and discusses the various roles of institutions. Chapter 8 provides systematic logic as to how the linkage of games across different domains can give rise to new forms of institutions, as well as the multiplicity of (suboptimal) institutional arrangements. The models of repeated games or evolutionary games applied in part I facilitate a rigorous analysis of individual institutions. However, because of the intended logical rigor, these models are also limited in their ability to capture the essential aspect of novelty, or innovation, in institutional change. Chapter 9 drops the assumption of objective fixation of the agents' action sets and introduces the concept of individual subjective game models through which agents subjectively view the structure of the games they play. By discussing how the agents cognitively revise their own subjective game models in response to external shocks or internal crises in a correlated manner, it attempts to describe a possible mechanism of institutional change. Chapter 10 turns to the objective mechanism of institutional change and discusses the diachronic interdependencies of institutions, leading to the path dependency of institutional change.

With a conceptual framework in place for analyzing the interrelationships among institutions across domains and over time, part III engages in comparative analyses of the practically relevant, and thus more complex, institutional arrangements of contemporary economies. Chapter 11 identifies several types of corporate governance institutions corresponding to different types of organizational architecture and examines their possible complementarities with institutions in other domains. Chapter 12 provides a new definition of relational financing and argues that despite the increasing globalization of financial markets, some types of relational financing based on the use of uncodified information may remain economically valuable. Chapter 13 deals with a case study to which the analytical and conceptual framework dealing with synchronic and diachronic institutional issues is systematically applied. It describes the mechanism of institutional emergence, coherence, and crisis with respect to a quintessential example of relational financing: the Japanese main bank system. Chapter 14 examines the Silicon Valley model and discusses under what conditions and in what sense this model can be an institutional innovation in the governance of technological product-system innovation. Chapter 15 concludes the book. It takes the analytical results developed along the way, and first identifies several important models of overall institutional arrangements and then present conjectures regarding why global, overall institutional arrangements will remain diverse despite increasing global integration of markets and the development of communications and information technology.