What are Institutions?

How Should We Approach Them? *

Masahiko Aoki

In 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 choose 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 1759*

The objective of this book is to develop a basic conceptual and analytical framework for understanding issues related to economic institutions such as: how do they evolve; why are their overall arrangements robust and diverse across economies; what impacts do they have on the economic performance of respective economies; and why (and how) do they or don’t they change in response to
changing environments (such as technological progress, global integration of markets, or demographic change)?

Economists have traditionally been engaged in analyzing the workings and implications of the market mechanism. Undoubtedly, markets can be considered one of the most salient institutions that human beings have ever produced. However, recently it has been increasingly recognized that institutions matter in order to understand the diverse economic performances of different economies and, when the phrase is cited, the reference does not seem to be limited to markets. Indeed, the last decade of the twentieth century witnessed various institution-relevant events and phenomena that have had, and in many cases will continue to have, significant impacts on the economic performance of the relevant economies and beyond. For example, think of the demise of the communist states and the subsequent transformation of their economic systems, the emergence of the Silicon Valley phenomena 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 impacts of global, non-governmental
organizations. Some of these may be thought of on the surface 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 adopted 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. We were told that there are two types of rules of the game: formal ones (constitutional, property-rights rules, and contracts) and informal ones (norms and customs). Then, even if good formal rules are borrowed from abroad, since indigenous, informal rules are inert and difficult to change, a tension may be created between the two. As a result, a borrowed institution may be neither enforceable nor functional.

Thus, economists became 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 that he or she is supposed to enforce? In short, how can the enforcer be enforced to do a proper job? A way out of the infinite regression in this reasoning may be to show that the rules of the game are endogenously generated, and thus become self-enforcing through the strategic interactions of the agents including the enforcer.

The most reasonable way of approaching institutions from this perspective is then to conceptualize an institution as an equilibrium outcome of the game. Thus, we have recently seen the publication of some important works based on views of an institution as an equilibrium of the 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 various institutions? Is it merely a bundle of more or less autonomous institutions, or does it exist as an internally coherent whole, i.e., as an equilibrium phenomenon of some sort?

Needless to say, viewing institutions (and possibly their complexes) as equilibrium phenomena 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? Theoretically, 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 it programmed by cultural genes? Can it be engineered by political entrepreneurs or engendered by mutant entrepreneurs? Do unforseen political events have stochastic impacts on the selection of an institution? Or, is there something more involved in the process of institutional change? Especially, how does the novelty that is often observed in the emergence of a new institution come about? These are some of salient issues related to institutional change/evolution.

Thus, the basic research agendas mentioned at the beginning of this introduction may be more specifically summarized in the following two problems: First, to understand the complexity and diversity of overall institutional arrangements across the (contemporary) economies as an instance of multiple
equilibria of some kind (let us call this the synchronic problem); second, 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 (let us call this the diachronic problem).

An understanding of the institutional diversity and complexity of (contemporary) economies requires an inquiry into the nature of the interdependencies of institutions across economic, political, organizational, and social domains, as well as the institutions linking those domains. In such an inquiry, it will be helpful not only to think within the framework of traditional economics but also to learn from some important contributions to institutional issues in neighboring disciplines, such as sociology, political science and the cognitive sciences. However, differing from the old institutional economics, we try to analyze the sources and implications of institutional diversity within a unified, generic game-theoretic framework rather than merely compiling a rich institutional catalogue 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 be useful for gaining a deeper theoretical understanding of the workings of the economic institutions.

However, we also need to be also aware that the game-theoretic analysis
alone in the traditional sense cannot be complete by itself as a systemic study of institutions. The analysis of the interdependencies of institutions within such a framework will indicate the possibility of multiple, sub-optimal, 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 mutually linked through the same markets. Thus, in order to understand why particular institutional arrangements has evolved in one economy but not in others, we need to rely on comparative and historical information as well. The implication is that institutional analysis must be also comparative and historical, thus it is called Comparative Institutional Analysis (CIA).¹

In considering the diachronic process of institutional evolution we will make an important departure from traditional game theory (either classical or evolutionary) at a later stage of the book (chapter 9). We will abandon the assumption that the players of the game (and its analysts) have complete knowledge of the objective structure of the game they play. Instead, they are supposed to have individual, incomplete cognitive views regarding the structure of the game they play—what we will 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 i.e., 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 cognitive crisis is created, their search for new subjective models may be triggered and continue until a new equilibrium is achieved. Thus, 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 overviews three different conceptualizations of institutions which economists have proposed. Section 2 tentatively introduces our conceptualization of institutions based on an equilibrium view, subject to a precise formulation later (chapter 7), and quotes five reasons why it is pertinent to our analytical purposes.
Section 3 introduces basic notions such as the game form and various types of domains that will play important roles in this book and finally previews the plan of the book.

1.1. Three Views of Institutions in Game-Theoretic Perspective

As mentioned, it seems that everyone now agrees that institutions matter. But such a statement does not make much sense unless we have a common understanding about what institutions are and how they are formed. 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. Leaving aside the old school of institutional economists, mainstream economists of the past were indeed engrossed with market analysis. But not only economics can make significant contributions to understanding the nature, origin, roles and implications of institutions, but also important economic phenomena and problems cannot be well understood without an analysis of institutions. Thus, an increasing number of economists have recently taken up the task of conceptualizing and analyzing institutions. As we will see, there are at least three different (although inter-related) 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 indispensable component of Comparative Institutional Analysis. The game analytic apparatuses we will apply in this book to deal with the synchronic problem, i.e., those borrowed from the theory of evolutionary and repeated games, are of relatively recent vintage. 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 interact with each other according to their own motivations, which 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 the game, the rules of the game, or equilibrium strategies of the players in the game.

When people casually talk about institutions in daily conversation, they
normally mean (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). However, North argues for a second view: that institutions should be identified with the rules of the game as distinct from its players. 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)

Such constraints may be informal ones (e.g., social norms, conventions, moral codes), or formal ones that are 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, but need to have been determined prior to playing the game. Since we are concerned with the origin of institutions, a question then immediately 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 will then be 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."(North 1995: 23)

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 can be expressed by specifying 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 pair of specifications a "mechanism" or a "game form." For example, consider the
mechanism of price control in which a seller is constrained by a ceiling which is set by the government on the sales price that he or she can charge. In this case the constraint on his/her choice set may be represented by a specific parameter value, i.e., the ceiling price.⁷ According to Hurwicz, however, further restrictions are needed 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 when 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 which 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, resource endowments). A social goal (such as efficiency, equity, clean air and water) 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 may always be sellers who may 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 which 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, which would then 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, equilibrium-of-the-game notion of institutions. One of earliest proponents of this third view is Schotter (1981).\textsuperscript{8} More recently, there have been two major developments in the equilibrium-of-the-game view of an institution based on different equilibrium notions: the evolutionary game approach and the repeated game approach. Representative works of the former approach are by Sugden (1986, 1989), Aoki (1995/2000), P. Young (1998), Okazaki and Okuno-Fujiwara (1998) and Bowles (2000).\textsuperscript{9} In the evolutionary game approach, a convention of behavior establishes itself without third party enforcement or conscious design. As a convention evolves, agents may 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 mutants and mistakes. Also, an articulation in words of conventionalized rules of conduct may be helpful in giving a clear answer in 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 has been 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 prisoners’ 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. 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 sub-game 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 which 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 were selected in the game.

The point may be illustrated by using the model of the merchant guild provided by Greif, Milgrom and Weingast (1994). This game was played repeatedly between multiple 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 needed to be organized in ways that secured 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 was established the ruler might be tempted to renege on the pledge. Suppose however that the merchants organize themselves into a guild and 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 before. Otherwise, they organize a boycott (let us leave aside the problem of the guild’s ability to enforce compliance from its members). The ruler does not cheat unless a boycott is announced by the guild. Once a boycott is announced, he 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 as a device to enable the ruler to credibly commit to his pledge and thereby to encourage trade expansion. In this example, the guild (an organization of the merchants) 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 that 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. phrases in [ ] are
Here, organizations refer to social entities, such as the guild above, that are a part of the set of players of the game and subject to constraints implied by an established equilibrium of the game. In this specific sense, this conceptualization may be said to incorporate 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 either by legislators, political entrepreneurs, or mechanism design economists. Among the equilibrium-of-the-game theorists, in the beginning it appears that there is no clear consensus on this issue. Those who adopt the evolutionary game approach clearly subscribe 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 feedback mechanism between their own and others' choices. How can individuals jointly select strategies that are mutually consistent and lead to the construction of an institution, especially when there are multiple equilibria? 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. For example, in the above parable of medieval trade, the combination of no-trade and cheating in each period (and thus the observation of only no trade) is another subgame perfect equilibrium. It seems natural then to consider that even those who adopt the super-rationality 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 between the players’ beliefs regarding the emergent situation, on the one hand, and the actual situation that is to be created by the choices of the players based on these beliefs, on the other? 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.


(A) 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. Although it is hard to provide a brief phrase for describing our definition and its full implications at the outset, we tentatively characterize it as an attempt to understand the nature of an 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 do regard them 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. Although we regard that an institution exists as beliefs of the agents, in order for these 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). Namely, 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. However, we keep the equilibrium notion behind our definition unspecified at this stage except that it is a Nash i.e., 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, as well as their liberal modifications. Also note that 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.

Consider the 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 pay-off 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) which specifies a pay-off distribution for each action profile from the domain as an exogenous rules of the game. Given these rules 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, and thus impact on 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 (see Figure 1.1. The dashed-line box represents an institution. Ignore the dotted-line box momentarily). In this way, although
endogenously created, an institution becomes objectified. 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- vs. exogenous-rules-of-the-game views of institutions.

Institutions as a self-sustaining system of shared beliefs may take various forms depending on the nature of the game (and thus a correspondingly appropriate concept of equilibrium), as well as ways in which an equilibrium is attained. An example is given by the afore-mentioned Greif-Milgrom-Weingast parable: If the ruler cheats merchants, the merchants will boycott his castle town for ever. This appears to prescribe an action-choice rule for the merchants to follow in the event of a ruler's cheating. However, as we have already discussed in the previous section, such an event may not actually happen, because a ruler's belief about the merchants' rational action-choice rule prevents him from taking that action. Therefore, the rule above can be considered as representing a belief of the ruler as well as that of the merchants. This is an example of regulatory
institution in which a credible belief of a coercive sanction constrains the actual action choice of agents. As suggested, institutions of this type may be analyzed as using the concept of subgame perfect equilibrium of repeated games.

On the other hand, since Durkheim sociologists have traditionally focused on normative expectations that morally prescribe certain action choices for all agents in a domain or a particular type of agent. However, norms may also be observed by agents because of their beliefs about social sanctions against default of prescribed obligations (such as social ostracization). Thus, they may be susceptible to game-theoretic analysis in a way that is similar to regulatory institutions. There are also rules and conventions that are actually constructed by the interactive or mimetic behavior of agents, but simply recognized as taken for granted. (Berger and Luckman 1966) In these cases, too, defection from conventions may be expected to be penalized by utility loss. Such mimetic mechanisms can be analyzed using the evolutionary game approach (chapters 2.1, 4.3). In reality, regulatory, normative and cognitive aspects may coexist in any institution (Scott 1995) and the choice of an appropriate equilibrium concept may partially depend on the level of abstraction/specification of the model that a researcher adopts (chapter 7.1 has more on this).

A salient feature of an equilibrium state recognized as an institution may

27
be represented in some explicit, codified symbolic forms, such as statutory laws, agreements, social structures or organizations as systemic arrangements of differentiated roles, and so on. However, the point is that a certain representation is an institution only if the agents believe in it. 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 shaken.

The equilibrium-summary-representation view of institutions helps clarify their dualistic constraining/enabling nature. The role of institutions is normally understood as a non-technological 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. 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, regardless of whether shared beliefs are regulative, normative or cognitive. However, an institution coordinates agents’ beliefs only as a summary representation of an equilibrium state. 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.

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 (marginal rate of substitutions in consumption and transformation) 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 numeraire (Koopmans 1957: Hurwicz 1960, 1973). Of course, in actuality markets are far

29
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).

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. 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.

(B) 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 summary-equilibrium-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-summary-equilibrium-representation view that emphasizes the cognitive aspect of institutions.¹⁹

(i) The 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 meta game 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 meta game and how do they play the meta-game? Where are the rules of the meta game determined? Thus, a problem of infinite
regression seems 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 either economic, social, or political exchange domains, while leaving the non-technological 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, i.e., 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 could be completely specified by technology. Even if it is possible to do so, however, there will possibly be multiple equilibria in a repeated game situation and, as already pointed out, which equilibrium is chosen from them cannot be determined endogenously. We need to consult historical events and rules prevailing in the
past, as well as rules prevailing in surrounding domains. It is so, because the particular subsets of actions which agents perceive as the sets of viable options may be constrained by historical precedents, while the way in which the consequences of a certain profile of agents' actions are determined in one domain may be 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 meta-game.

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 still another enforcer to enforce the rules of action prescribed for the original enforcer? As the preceding discussion regarding Hurwicz's contribution suggests, a solution to this problem is again to
analyze a game including the enforcer, if any, as a player, and 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, however, the presence of the enforcer may need to be presumed as a player of the game at the outset, presumably given by history.

(ii) *History matters* : Second, by showing the possibility of multiple equilibria in specific models, the institution-as-an-equilibrium approach is able to 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. 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, which may 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, yet 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—the phenomenon 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: (1) based on historical and comparative information, first sort out what technological and institutional factors can be treated as \textit{exogenous} and what institutional factors are to be treated as \textit{endogenous}, i.e., to be explained; (2) 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; (3) next, find out if some of these solutions are useful for understanding the nature of the institutional factors needing to be explained; and (4) finally examine what historical factors can be considered responsible for the selection of that particular equilibrium solution to determine the role of history.

\textit{(iii) Inter-linkages and inter-dependency of institutions:} Third, 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 a so-far non-existent institution, their implementation in particular economic, political and social contexts can often have unintended consequences. For example, even if a government in a post-Communism economy 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. This situation is somewhat analogous to the one in which a medicine which has been tested in a laboratory may have unpredicted side-effects when it is administered to a human being because of the complexity of the living organic system. A major reason for such unintended outcomes is the absence of fits between the designed plan and the existing institutional environments which reflect 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 life-time 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, 12).

We will conceptualize possible institutional interdependencies as
institutionalized linkage and institutional complementarities in Part II. These intuitively appealing notions can become amenable to rigorous analysis when the equilibrium-oriented notion of institutions is applied. Specifically, we consider games in different domains of the economy, such as organizational coordination, commodity trade, transactions of services of human and financial assets, political-transactions, social-exchange, etc. Then, applying 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 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, as well as 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. 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.
(iv) Institutional change through the competition in a market of symbolic systems:

Fourth, the summary-equilibrium-representation view of an institution may suggest a new way to approach the mechanisms of institutional change. As noted already, the ability of an institution to transmit information is not complete. But for the bounded-rational individual agents the compressed information may be adequate. They can be guided by it to develop skills or dispositions that are consistent with the endogenous rules of the game. However, when patterned choices become problematic because of environmental and internal changes, an institutional crisis in the cognitive sense may be triggered: the taken-for-grantedness of an institution, i.e., shared beliefs regarding the ways in which the game is played, begins to be questioned and the agents are 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 this 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, characterized by a move from an equilibrium under given sets of action-choice rules of agents to an equilibrium under other (possibly
expanded) 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 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 serve as their summary representation of equilibrium incorporated into agents stable beliefs, i.e., as an institution (see the line from the dotted box to the dashed box in Figure 1.1). The point is that some symbolic system 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 in some cases 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.

(v) The role of statutory laws and public policy discourses: Finally, but not least important, the difference in whether the rules of the game constituting institutions are viewed as endogenous to the relevant domain or 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 yet that they matter to the performance of an economy, its implications could then be that a badly performing economy should reform itself by designing and implementing better rules, possibly emulating best practices.
elsewhere. If this was not realized, blame could then be placed on politicians. However, there are two problems with these kinds of arguments. 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.\textsuperscript{26} Blaming the low morality or incompetence of politicians alone may not help resolve a real problem.

Secondly, as already noted, a policy may not yield the outcome intended by the government or politicians if it does not have fits 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.\textsuperscript{27} Statutory laws or regulations may induce the evolution of an institution but they themselves are not institutions. It may be also the case that an institutional outcome will be different from what a legislature or government initially intended. A careful and systematic study is called for regarding such issues as: 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 policy impacts on 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.\textsuperscript{28}

A subtlety of the issue discussed now is that endogenizing the government does not necessarily mean, however, that the outcome of the game we are studying will become fully determined, leaving no scope for policy advising the paradox referred to as the determinancy paradox (Bhagwatti, Brecher and Srinivasan 1984). As hinted previously, when a certain domain is stricken by 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 can be latitude for some exogenous symbolic systems to compete for the position of an attractor or a focal point (Schelling 1960) for the formation of coordinated beliefs. These could well be programs or platforms of competing political parties, professional political advice, an elites pact (Weingast 1997) or drafting and enactment of statutory laws. Thus, political discourses in and out of polity may have a certain imprint on subsequent institutional evolution.\textsuperscript{29} The point is that whatever they may be, one of the crucial factors that will determine their impacts is their fits with emergent
practices in domains other than the polity.

1.3. Organization of the Book

(A) Domains, Game Forms, and Institutions

We have introduced the basic conceptualization of institutions to be further developed in this book. We plan to apply this concept and examine its implications to some of salient contemporary and historical institutional issues from a comparative perspective. For doing so, we need to develop a unified framework for systematically analyzing the interdependencies of various institutions clustering in various economies. Also, it will be worthwhile and helpful if we can develop a unified conceptual and analytical framework that can incorporate contributions and insights into institutional issues in economics, as well as sociology, political science, and the cognitive sciences, whenever possible. In order to develop such framework, in this section we first make precise basic terms and concepts, such as domains and the game form, that will play an important role in our framework, and identify several basic types of domains. It finally present the plan of the book.

We will make a domain of the game as a unit of analysis. A domain of the
game is composed of a set of agents, either natural persons or organizations, and sets of physically feasible actions open to each agent in successive periods.\textsuperscript{30} Recall that a combination of actions chosen in one period by all the agents in the domain is termed as an action profile. An action profile determines the distribution of the payoffs among the agents in the domain. We decompose the pay-off functions 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 pay-off 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, which are to be distinguished from an institution as the endogenous rules of the game. Also, in games that we will deal with in his 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 the domain choose their private plans of action so as to maximize their current payoffs, or the present value sum of their current and future pay-offs, 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. The choice of an action plan thus chosen by an agent is his/her strategy. A strategy may prescribe only one particular action, or a comprehensive plan (script) 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 and polity domains, as well as generic,
organizational fields. We investigate what types of institution/convention can become viable on and across those domains. As mentioned in section 2, North has excluded organizations from the category of institutions. He contends that organizations are players of the game rather than the rules. His reason for doing so is to pinpoint the role of organizations as agents of institutional change in the polity domain. In contrast, this book will place relatively less emphasis on politically-determined formal rules, as well as on the causality from the polity to the economic process, although mutual feedback mechanisms between the outcomes of games in the polity domain and those in the economic-transaction domain are admittedly important and we try to deal with them as such. On the other hand, this book will place substantial emphasis on organizational conventions of business enterprises and other private-order entities (such as 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, as well as variations of the choice sets across the agents. We try to do so in terms of their technological properties as much as we can, in order to identify variations in institutions that may endogenously evolve in each domain or across domains. However, as already
pointed out, it is, in principle, impossible to start our analytical discourse on institutions in a purely institution-free, technology-only manner. 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 own pay-offs 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 dis-economies throughout the domain (e.g., congestion in the use of common resources, free-riding in maintenance efforts, or benefitting others by one’s own effort to produce common resources), but the agents may not be able either to exit from the domain nor to be expelled from it. An endogenous institution that may arise in this domain could be (customary) property-rights rules, 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 which they can dispose of or 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, etc. 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, 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. Trade domains and organizational domains

51
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 is somewhere between the trade and organization domains, but distinguished from both. 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
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, industrial associations and unions, economic classes, and so forth, depending on 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, 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, arrest) as well as the monopolized supply of public services such as law enforcement. Private agents cannot escape from the impacts of government actions by choice, and thus the latter has exclusive compulsory regulatory power. However, private agents may choose to support or not support (or 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 is the (sociological) domain that will not be dealt with on its own in this book, but will play an important complementarity role in understanding some institutions, such as community norms and status differentiation or rank hierarchies within homogenous teams or organizations.³⁵ In this domain, non-economic goods/bads (social symbols, languages) which would directly affect the pay-offs 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 depicts the demarcation of the six types of domains of games in terms of two dimensions. 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 the above characterization.

As mentioned, any classification of domain types cannot be made in purely technological terms. As a result, the demarcation of domains along the lines presented above would be more or less ad hoc. If we take a particular set of agents who are strategically interacting with each other, classification may not neatly fall in with one of the types, but have multiple domain characteristics. For example, consider the firm. Needless to say, its most salient characteristics are organizational in incorporating the 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 domain. Generically, its essence can also be captured as an outcome of matching agents 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). As already mentioned, it has a modicum of the aspect of a commons domain as well (e.g., in the use of intangible information assets). Finally, but not least important, the firm is considered to be embedded in some governance structure that resembles an institution in the polity domain (e.g., corporate firms, 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 at one time. However, our ultimate objective is to understand the logical structure in which characteristics of the various facets of a business firm fit each other in alternative ways, depending on historical and environmental contexts.

Finally, let us tentatively reiterate the intuitive concept of institutions based on the shared-belief cum summary-equilibrium-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 could be another
equilibrium (or more generally, another sequence of equilibria), we identify an institution as follows:

An institution is a self-sustaining system of collectively shared beliefs about non-technological, self-enforcing rules of the game that govern the strategic interactions of the agents. Its content 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, and in turn reproduced by their actual choices in a continually changing environment.

Five elements are present in this conceptualization: endogenicity (implied by self-sustaining, self-enforcing, and reproduced), information compression (implied by a compressed representation), robustness with respect to continual environmental change and minor deviance (implied by invariant features of an equilibrium path, perceived by almost all the agents and reproduced ..., in a continually changing environment), universality of relevance (implied by collectively shared, 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. For example, 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 exchange 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 a myriad of 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 could be possible in each of these domains. Not only that, but institutions can also 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, in spite of complexity and variety, they may not be randomly arranged. One major purpose of this book is to discover the generic laws of regularities that prevail in various overall institutional arrangements.

(B) The Plan of the Book

This book is composed of three Parts. Part I takes up successively the six types of domains in primitive forms and identifies the prototypes of institutions as stable multiple equilibria of games in each domain. This provides a foundation for a generic, game-theoretic framework to be built in Part II for conceptualizing institutions, analyzing their interdependencies across domains and over time, and thereby understanding the mechanism of institutional change. Then, we need to check if this framework actually be 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 that 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 (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 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, without an explicit analysis of the inter-dependencies among them, and thus in an inevitably primitive form. This Part is largely reliant on the previous works of many authors, including myself, and provides a rough overview of the current state of analytical approaches to institutions (it is by no means intended to be a substantial comprehensive survey, however).

Building on the preparatory, taxonomic analysis of Part I, Part II engages in the construction of a generic, analytical framework for institutional analysis,
with illustrative examples and cases. The basic motive behind this is to provide a unified framework within which the systemic nature of overall institutional arrangements of the economy, as well as their changes, can be analyzed. Chapter 7 provides a precise game-theoretic conceptualization of institutions as a self-sustaining system of shared beliefs along with being a 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 (sub-optimal) institutional arrangements. The models of repeated games or evolutionary games applied in Part I facilitate a rigorous analysis of individual institutions. However, precisely because of the intended logical rigor, these models are also limited in their ability to capture an essential aspect of institutional change: novelty or innovation. 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.

Having laid down a conceptual framework 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, in spite of the increasing globalization of financial markets, some types of relational financing based on the use of uncodifiable information may remain economically valuable. Chapter 13 deals with a case, the Japanese main bank system, 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 constructs 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. The book concludes with chapter 15. Based on analytical results developed along the way, it will first identify several important models of overall institutional arrangements and then present conjectures regarding why global, overall, institutional arrangements will remain diverse in spite of the increasing global integration of markets and the development of communications and information technology.
Figure 1.1. An Institution as Shared Beliefs Formed as Summary Representations of an Equilibrium

Individual Agents               Domain

An institution is represented by the broken-lined box.
**Figure 1.2. Six Types of Domains of Games**

<table>
<thead>
<tr>
<th>set of agent</th>
<th>variable</th>
<th>fixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>set of choices</td>
<td>symmetric</td>
<td>organizational file</td>
</tr>
<tr>
<td></td>
<td>asymmetric</td>
<td>trade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>social exchange</td>
</tr>
<tr>
<td></td>
<td></td>
<td>organization</td>
</tr>
</tbody>
</table>
Endnotes


2. For Durkheim, social institutions are composed of symbolic systems of knowledge, beliefs, and the collective ideas. These systems are a joint product of human interactions, but they are experienced by individuals as objective and coercive (Durkheim, 1901[1950]).

3. The definition of institutions by Veblen was settled habits of thought common to the generality of man. (1909:239)

4. Ostrom is another notable advocate of this view (1990: 51).

5. Although enforcement of formal rules is made exogenous to the game by a third party, the performance of the game would depend upon the effectiveness of monitoring and sanctions. North asks: "What happens when a common set of rules is imposed on two different societies? .....The results, however, are not
similar .....Although the rules are the same, the enforcement mechanisms, the way enforcement occurs, the norms of behavior, and the subjective models of the actors are not. Hence, both the real incentive structures and the perceived consequences of policies will differ as well. Thus, a common set of fundamental changes in relative prices or the common imposition of a set of rules will lead to widely divergent outcomes in societies with different institutional arrangements. (North, 1990:101). He therefore invokes a more comprehensive notion of an "institutional framework," which includes "legal rules, organizational forms, enforcement, and norms of behavior"(North 1991: 33. See also 58, 101. Elsewhere he uses the expression institutional scaffolding (North, 1998) to refer to a similar set).

6. The range of the outcome function is the space of physical outcomes rather than that of utility pay-offs. Hurwicz considers that players' utility functions are not part of the rules of the game. Thus he uses the word game-form, rather than game, to refer to the trio of the set of players, their choice spaces, and the outcome function.

7. Strictly speaking, such a parametric representation may be too restrictive and narrow for identifying the institution of a price ceiling, however, as the
government may have discretion on the actual level of the price ceiling. Therefore, Hurwicz allows room for variations in the parameter value and derives the concept of an institution as a family of mechanisms (e.g., a range of ceiling prices) rather than a single mechanism (e.g., a particular ceiling price).

8. Schotter defined a social institution 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." (1981:11) In this conceptualization, as in the second view, institutions are regarded as rules of conduct. However, instead of being given exogenously by the political or legislative process, such rules are assumed to be endogenously created within the economic process as a solution to the game.

9. A general overview of evolutionary thinking on institutional economics is given by Hodgson (1993) although it does not adopt an explicitly game-theoretic framework.

10. A sociological equivalence of a comprehensive plan of actions may be found in the notion of script introduced by Schank and Aberson (1977), which describes behavioral patterns and sequences called up by specific roles or situations.
11. For Greif’s most recent conceptualization and its ramification, see Greif (1998b). One subtle difference between his and our conceptualization that will be made clearer in the discourse of the book is our explicit treatment of the cognitive mechanism in the process of institutionalization and its implication to institutional evolution (see, particularly chapters 5.2, 7.2, 9.2).

12. We will argue in chapter 7 that alternative notions of equilibrium, particularly classical vs evolutionary, are not necessarily mutually exclusive, but may be complementary. Also, depending on the level of specification of domain characteristics, an appropriate equilibrium concept may differ.

13. This aspect is consistent with the phenomenological view of institutions initially developed in sociology of knowledge by Berger and Luckmann (1966). According to them institutionalization occurs whenever there is a reciprocal typification of habitualized actions by types of actors. (1966:54) But this institutional world is experienced as an objective reality. (ibid.:60)

14. On this point from sociological perspective, see Coleman (1990), particularly Part II.
15. Thus, Searle, the author of *The Construction of Social Reality*, states that all institutional facts are ... ontologically subjective, even though in general they are epistemically objective. (1995:63). However, also note that Berger and Luckmann who founded the phenomenological approach to institutions in *The Social Construction of Reality* submitted that (the institutional world) does not ... acquire an ontological status apart from the human activity that produce it. (1966: 60) Our equilibrium-oriented approach admits that both outcome of human activities and beliefs are mutually constitutive as institutions (see figure 1.1).

16. As an example, we will see in Chapter 12.2 that the Japanese main banks system was an institution composed of a particular system of beliefs, but not by explicit statutory laws, and how that belief system came to be demised through the strategic interactions of the players in financial and corporate domains under changing international and technological environments.

17. Arrow submits: "Expectation *per se* can be thought of as an element of individual psychology, but in practice social institutions play a major role in guiding and forming expectations. There are . . . understandings that others will not exploit every possible short-term profit opportunity, and elaborate financial services networks to provide forecasts and to smooth out temporary
difficulties." (Arrow 1997:6)

18. Later I will make this robustness property as one requirement of an institution. In that sense, the price vector as a summary representation of tastes and technology cannot be an institution, as they adapt in response to changes in these parameters. Therefore, the analogy stops here. However, beliefs that one can buy goods by paying agreed-on prices and have them delivered as promised can be an institution.

19. The following points have evolved from my own writings in the past such as Aoki (1994a, 1995/2000, 1996). Also Greif has developed the same ideas based on his original research (Greif 1989, 1994, 1998b) and presented them in (Greif, 1997, 1998a) in succinct forms. His influence on my thinking is clear from the following statement. Also, contributions to the third point by Milgrom and Roberts (1990a, 1992, 1994) and my indebtedness to them need to be noted.

20. For the latter approach, see Reiter and Hugh (1981) and Hurwicz (1996). At one point Young adopts this approach from the evolutionary game perspective (1998:chapter 9).
21. Basu (1997b) provides an interesting parable indicating this point. Compare two games. One of these games call it the law game has law and the public officials, while the other call it the anarchic game does not and thus is institution-free. However, suppose that these two games do not differ in terms of the available strategy sets or payoffs to the agents and in any other respects because any action available to an agent labeled the public official in the former is available to the same agent (without that name) in the anarchic game. Suppose that we can identify the equilibrium in the former lawful equilibrium and another as anarchic equilibrium. Then it may appear that legal rules are simply equilibrium. However, the question remains as to what factor it would be to make the lawful equilibrium chosen in the former rather than the anarchic equilibrium. There must be some extra-technological aspect of situation that has some effect on the choice of the equilibrium. In this case, the law constitutes such a focal point (Shelling 1960).

22. The multiplicity of equilibrium as an requirement of a convention is emphasized by Sugden. He submits that "[a] self-enforcing rule ...could be regarded as a convention if and only if we can conceive of some different rule that could also be self-enforcing, provided it once become established" (Sugden,
23. This is discussed in more detail in example B1 of chapter 10.3(B).

24. For own contributions of Milgrom and Roberts to the comparative analysis of institutions and organizations, see Milgrom and Roberts (1990b, 1992, 1994), as well as Holmstrom and Milgrom (1994). Also Pagano (1993) and Pagano and Rowthorn (1994) are one of the earliest analytical contributions to institutional complementarity.

25. Transaction cost economics regards the transaction as the basic unit of analysis as opposed to the agency theory which regards the individual agents as such (see Williamson 1996, chapter 7 for the comparison of the two approaches). In that sense the approach of transaction cost economics is similar to ours which regards the domain of the game as the unit of analysis. Williamson, an authority of transaction cost economics, makes explicit about the institutional nature of the environment of the unit of analysis under his study: [t]hese definitions of institutions [such as by North, Schotter, and others] mainly operate at the level of the institutional environment, the so-called rules of the game. The second, more microanalytic, level at which institutional economics works is at the level of
institutions of governance. This book is principally concerned with the institutions of governance (markets, hybrids, hierarchies, bureaus). (Williamson 1996: 4-5).


30. Later in chapter 9 we deal with the case in which agent’s subjective perception defines his/her activated set of action choices.

31. Thus the notion of common resources here is close to that of common-pool resources used by Ostrom. She, however, distinguishes the latter from public goods (which are provided by the public authority), the services of which are subjected to joint uses (Ostrom 1990:30-3). We do not necessarily exclude
collective goods/bads, such as environmental pollution, from (negative) common resources, as long as the sets of actions are qualitatively symmetrical across the agents in the domain. When the common resources are unilaterally provided by the public authority, we call them public goods and relegate their analysis to the polity domain discussed below.

32. We are primarily concerned with the coordinating aspect of the organization here, although the organization may have other aspects as mentioned momentarily. From this perspective, even organizations that are egalitarian or collectivist from the governance perspective may be viewed as having the focal agent (the management).

33. This is a defining characteristic of organizations according to a classical treatise on organizations by Barnard (1938).

34. By distinguishing the government as an autonomous player of the game, our approach in this book is in accordance with lines of thought developed by political scientists such as Skocpol (1979, 1985), Krasner (1978, 1984), and Weingast (1997). However these authors refer to governments (sovereigns) by state rather than as an equilibrium state in the polity. For various concepts and
approaches to state in political science, see Krasner (1984). For game theoretic literature see endnote 26 and 28 above.

35. This type of domain has been the objects of institutional study by diverse schools of sociologists, such as the social exchange theorists (e.g., Blau 1964, Homans 1961), phenomenological theorists (e.g., Berger and Luckman 1966), new institutional sociologists (e.g., Meyer and Rowan 1977, Zucker 1977, Meyer and Scott 1983, DiMaggio and Powel 1983), social embeddedness theorists (Granovetter 1985), and rational choice theorists (e.g., Coleman 1990). For a survey of sociological and other social scientific approaches to institutions, see Scott (1995).

36. In that the exchanges of symbols and languages directly affect the payoffs of the agents, they are distinguished from the so-called cheap talk in game theory.