Chapter 5 Endogenous Institutions and Game-theoretic Analysis

Chapters 3 and 4 illustrate that restricting the set of admissible institutionalized beliefs is central to the way in which game theory facilitates the study of endogenous institutions. Durkheim (1950 [1895], p. 45) recognized the centrality of institutionalized beliefs, arguing that institutions are "all the beliefs and modes of behavior instituted by the collectivity." But neither Durkheim nor his followers placed any analytic restrictions on what beliefs the collectivity could institute. Because beliefs are not directly observable, however, deductively restricting them, as game theory lets us do, is imperative. The only beliefs that can be instituted by the collectivity—that can be common knowledge—are those regarding equilibrium (self-enforcing) behavior. Furthermore, the behavior that these beliefs motivate should reproduce—not refute or erode—these beliefs.

Game theory thus enables us to place more of the "responsibility for social order on the individuals who are part of that order" (Crawford and Ostrom 1995, p. 583). Rather than assuming that individuals follow rules, it provides an analytical framework within which it is possible to study the way in which behavior is endogenously generated—how, through their interactions, individuals gain the information, ability, and motivation to follow particular rules of behavior. It allows us to examine, for example, who applies sanctions and rewards that motivate behavior, how those who are to apply them learn or decide which ones to apply, why they do not shirk this duty, and why offenders do not flee to avoid sanctions.

The empirical usefulness of the analytical framework of classical game theory is puzzling, however, because this theory rests on seemingly unrealistic assumptions about cognition, information and rationality. For example, the analysis requires a common knowledge of rationality and that the players have a complete and closed model of the situation. It is possible to respond to this puzzle by taking comfort in the empirical usefulness of the theory. It works. But it is beneficial to go beyond this position. Accordingly, this chapter asks what is revealed about endogenous institutions by the need to impose various unrealistic assumptions when studying behavior in strategic situations?

Similarly, the chapter asks what do we learn about institutions from the game-theoretic

insight that coordination problems are common? In strategic situations, each individual is better off playing the strategy that is the best response to the particular equilibrium strategies others are following. Yet game theory shows that multiple equilibria usually exist in the repeated situations central to institutional analysis. This multiplicity of equilibria implies that one will seek ways to coordinate his behavior because deduction alone—knowledge of the structure of the situation—is insufficient for finding one's best response.[1]

The aim of this chapter is to address these questions to better understand institutions and the extent to which game theory can be used to study them. Doing so requires examining the cognitive, coordinative, normative, and informational microfoundations of behavior, how institutions provide them, and how the implied behavior then reproduces these institutions. In conducting such an examination, the chapter draws particularly on learning and experimental game theory, cognitive science, and sociology.

Section 5.1 emphasizes the importance of socially articulated and disseminated rules in providing individuals with the cognitive, coordinative, and informational microfoundations of behavior. These social rules provide an individual with the information and informative and the cognitive model (also referred to as mental models or internalized belief system) required to choose behavior. (Henceforth I will use the terms cognitive model, mental model, and internalized belief system interchangeably.) Similarly, social rules coordinate behavior by providing a public signal regarding the behavior that is expected of individuals in various circumstances. In short, social rules constitute the heuristics that enable and guide behavior by helping individuals to form beliefs about the world around them and what to expect from others.

Commonly known social rules enable and guide behavior, and retrospective individuals with limited rationality and cognition respond to these rules. On the one hand, each individual takes the cognitive, coordinative, and informational content of institutionalized rules as given; he responds to (or plays against) the rules, taking them as given. On the other hand, because each individual responds to these rules based on his private information and knowledge, such rules aggregate information and knowledge and distribute it in a compressed form.

---

[1] Indeed, coordination problem is a characteristic of every game that is not dominance-solvable.

The only social rules that can be institutionalized—that can be common knowledge, expected to be followed, and correspond to behavior—are rules that each individual finds optimal to follow, given his private information, knowledge, and preference. In situations in which institutions generate behavior, institutionalized rules and the associated beliefs correspond to self-enforcing behavior. Finally, because behavior corresponds to the institutionalized rules and the associated beliefs, these rules and beliefs are reproduced—not refuted—by behavior.

In situations in which institutions generate behavior, institutionalized rules, the corresponding internalized and behavioral beliefs, and the behavior that these beliefs motivate constitute an equilibrium. A structure made up of institutionalized rules and beliefs enables, guides, and motivates the self-enforcing behavior that reproduces it. Most individuals, most of the time, follow the behavior that is expected of them.

Section 5.2 employs this understanding of institutions to highlight why the game-theoretic analytical framework, which rests on seemingly unrealistic assumptions about cognition, information, and rationality, has been a useful tool for positive institutional analysis. Understanding why this is the case is essential to knowing when game theory can be usefully employed and to what extent.

The section argues that the game-theoretic analysis, which assumes a complete model and common knowledge and focuses on equilibrium strategies played by highly rational individuals, corresponds to a situation in which institutionalized rules that aggregate private knowledge and information provide shared cognition, information, and coordination. The game-theoretic analysis restricts the set of admissible rules, beliefs, and behavior to those in which each limitedly rational individual, responding to the cognitive, coordinative, and informational content of the institutionalized rules, follows the behavior expected of him.

In situations in which an institution generates behavior, the knowledge and information that are compressed into the institutionalized rules enable and guide individuals, despite their limited perception, knowledge, and computational ability, to act in a manner that leads to behavior and reflects the constraints on admissible beliefs and behavior that the game-theoretic equilibrium analysis captures. Classical game theory can be usefully employed to study situations in which it is reasonable to assert that social rules were institutionalized.

For simplicity of presentation, sections 5.1 and 5.2 ignore norms and social considerations. Section 5.3 supplements their discussions by elaborating on how to integrate normative and social considerations into the analysis. Indeed, a promising aspect of game theory is its ability to provide a unified analytical framework to study the cognitive, coordinative, normative, and informational foundations of behavior while capturing the response of individuals to both social, normative, and materialistic considerations.[2]

Distinguishing between the object of study (institutions) and the analytical framework used to study it (game theory) is central to this chapter. This distinction is also the focus of studying the dynamics of endogenous institutions, the topic of Part III. To lead into this topic, section 5.4 explains why it is appropriate to study institutions without examining their origins. It also notes that legitimacy is crucial to the institutionalization of intentionally created institutions. Different societies can and do have distinct norms regarding legitimacy, each entailing a distinct institutional development. The different sources of legitimacy that established themselves in late medieval Europe and the Islamic world still prevail today. Section 5.5 summarizes the chapter's argument and delineates directions for further development.

The argument made in this chapter rests on a particular notion of rationality, maintaining that when institutions generate behavior, socially articulated and disseminated rules span the domain that people understand and within which they can act rationally. At the same time, the chapter recognizes that individuals are motivated by social and moral considerations. Are these two premises consistent with each other? Is it appropriate to consider individuals as strategic while recognizing that social and normative considerations influence their behavior? Appendix B presents evidence supporting the claim that although individuals have social and normative propensities, they are nevertheless rational in the above sense.

Before proceeding, it is important to emphasize what this chapter *is not* about. The chapter is about institutions and not about their dynamics. It focuses on regularities of behavior in the population as a whole while ignoring forces and factors that direct particular individuals to

[2] The discussion in Aoki (2001, pp. 13-14, 235-9, 412-13) is closest to the one developed here. He argued that institutions provide "summary representation" of situations and that the set of summary representations is constrained by individuals' responses to them.

act differently thereby sometimes leading to institutional change. I return to this important issue in Chapter 6.

5.1 Institutionalized Rules, Institutions, and Equilibria

The behavioral choices an individual makes require both a cognitive model (also referred to as a *mental model* or *internalized belief* system) and a sufficient amount of the right information. (See Hayek 1937; Savage 1954; and North 2005). Cognitive models constitute one's understanding of the causal relationship between actions and outcomes. Although they are usually incomplete they underlie rational as well as habitual and mimetic behavior (see Denzau and North 1994; Eysenck and Keane 1995; Clark 1997a, 1997b; and Mantzavinos 2001).[3] In addition, however, behavioral decisions require appropriate information about the particularities of the situation. For example, if one's model asserts that sufficiently religious people can be trusted, acting on this premise requires knowing the extent of peoples' religiosity.

Classical game theory is mute regarding the sources of the cognition and information required for behavior. But the analysis requires a strong and unrealistic assumption about players' cognition and information, calling attention to how, whether, and to what extent this requirement is met in the real world. The analysis requires that players have a complete and closed model of the situation and correct common priors.[4] It assumes that it is common knowledge that each player has complete information about the details of the situation, including causal relationships, other players' preferences, and the magnitude of various parameters. When such information is missing, the players assign the correct prior probabilities to all possible values of the unknown parameters. Each player assumes that his opponents are rational, that they

---

[3] Indeed, it is part of human nature to seek a rationale for actions ex post and to try to explain and develop a cognitively coherent account of past experiences to guide future behavior.

[4] The discussion here counters the assertion that game theory is inappropriate for institutional analysis because it assumes that the rules of the game are common knowledge, although common knowledge is neither necessary nor sufficient for the Nash equilibrium condition to hold (see Aumann and Brandenburger 1995). Nash can prevail or be reached in evolutionary and learning games without common knowledge, while common knowledge is sufficient only for the weaker equilibrium notion of rationalizable equilibrium (Bernheim 1984; Pearce 1984), the essence of which is the iterated elimination of dominated strategies.

model the game exactly as he does, and that they assign the same correct priors. Even after making these assumptions, the computational complexity required to find an equilibrium is daunting, even in moderately complicated games. How, then, can we expect real-world actors to reach an equilibrium when, as is common in complex situations, they lack a complete model? How can we assert that individuals can rationally calculate their way through games that are difficult even for the modeler?

Behavioral choices in social situations also rest on a player's ability to coordinate his behavior with that of others. Whether it is best to drive on the right or the left depends on what others are doing. Even in simple, repeated strategic situations, such as the prisoners' dilemma game, multiple equilibria usually exist (see Appendix A). Because there are multiple equilibria and because the behavior that serves one best depends on the particular equilibrium behavior others are following, rationality alone is insufficient to guide one how to behave. One faces a coordination problem. In the case of the Maghribi traders, for example, the strategy calling for merchants not to hire agents and for agents to cheat is also an equilibrium. Multiple equilibria imply that ex ante deduction is insufficient for choosing behavior (see, e.g., Schelling 1960; D. Lewis 1969; and Sugden 1989). Yet, one seeks to know—because it is beneficial for one to know—what particular strategy is followed by others. How do individuals choose behavior given that, even in the simplified world captured in the game-theoretic models, rationality alone is insufficient for making choices?

To solve a game there is a need to impose restrictive assumptions, such as the assumptions that individuals are highly rational, that they have the same cognitive understanding of the situation, and that all this is common knowledge. The analysis reveals the importance of coordination, as it indicates that multiple equilibria exist in strategic repeated situations. What does the need to impose such assumptions tell us about the real world? How can we assert that an analytical framework based on such unrealistic assumptions is useful for positive analysis? How and to what extent are these assumptions met in the real world?

Economists responded to these challenges by exploring whether learning by individuals

with limited knowledge and information can lead to self-enforcing regularities of behaviors.[5] The theory of learning in games asks if a rule of behavior corresponding to a Nash equilibrium can reflect individualistic learning. It turns out that reaching Nash equilibrium requires replacing the very demanding assumptions of classical game theory with a set of other demanding and unrealistic assumptions.[6] Learning models often require that individuals be completely myopic, implying unreasonable behavior, such as not performing a costly experiment no matter how high the resulting expected return might be. These assumptions are very restrictive, but not imposing them makes the analysis too complicated to provide a convincing account of how individuals learn.

Focusing on individualistic learning, however, ignores the social context within which institutionalized behavior takes place. In this context, socially articulated, disseminated, and commonly known rules provide individuals with the cognitive, coordinative, and informational foundations of behavior. In order to act, each individual needs a cognitive framework, information, and the means to coordinate his behavior. Individuals seek these microfoundations of behavioral choices at the social level at which it is provided in the form of social rules. Sociologists have long noted that when taking actions, members of a society are aided by rules providing "socially sanctioned facts of life that any bona fide member of the society knows" (Garfinkel 1967, p. 76). Decision making at the individual level is done in the context of commonly known social rules that provide a cognitive system, information, and coordination.

These rules are shared by members of a society: everyone knows them, and everyone knows that others know them. The rules can emerge spontaneously (e.g., in the form of social norms) or deliberately (through a political process); they can be formulated quickly or over a long period of experimentation and social learning. Social rules are transmitted in diverse forms,

---

[5] These models generally focused on learning about others' strategies or various parameters of the models. The lack of a theory to account for the emergence of endogenous cognitive systems based on individualistic learning lends support to the assertion made later about the importance of social rules.

[6] Schotter's (1981) seminal work pioneered the application of learning game theory to institutional analysis. Regarding learning, see Marimon (1997); Fudenberg and Levin (1998); Rubinstein (1998); and Young (1998). Evolutionary game theory was another response to justifying the use of Nash equilibria. Chapter 1 argues that it suffers from drawbacks similar to learning game theory.

through laws, regulations, customs, taboos, conventional rules of behavior, and constitutions. They are articulated and disseminated by such socializing agents as parents, teachers, peers, priests, tribal elders, and CEOs; they become identical and commonly known during the socialization process, during which they are unified, maintained, and communicated. They are transmitted by myths, fables, holy scriptures, educational systems, public announcements, manuals, and ceremonies and disseminated by various carriers, such as parents, teachers, priests, and regulatory agencies.[7]

Such socially articulated and transmitted rules contain a cognitive system that embodies, transmits, and propagates knowledge and information reflecting the accumulated experience and innovativeness of past and present members of the society. A cognitive system provides the terms for describing socially recognized and created items, ideas, actors, events, and possible actions to which the system also imputes meanings. It articulates the objectives and capabilities of various actors and the outcomes associated with various circumstances; those who speak publically against a dictator suffer, for example, and those who are honest in per-period profitable exchange prosper. A cognitive system constitutes a shared cultural understanding (a script or interpretive frames) of the way the world works (Zucker 1983, 1991; Meyer and Rowen 1991; DiMaggio and Powell 1991a; Dobbin 1994; Scott 1995). It provides typification, classifications, and meanings, using symbols such as words and signs. In a sports game, for example, the cognitive system enables us to communicate and comprehend various physical items (such as a basketball), ideas (winning), events (fouls and free throws), actors (a captain and coach), and the set of events or actions that fall into a particular category (such as those that entail winning). (See D'Andrade 1984; Searle 1995; and Scott 1995.)

Using the typification, categorization, and cognition provided by the cognitive system, the "behavioral rules" component of social rules specify what is expected of individuals with particular social positions in various circumstances: members of the two basketball teams have to stand in particular positions during a free throw, a driver must stop at what is cognitively defined

[7] See, for example, K. Davis (1949, in particular pp. 52ff., 192ff.) and Bandura (1971). Even the form of circular seating in organizations for collective decision making—from the ancient Native American councils to the U.S. Congress—is aimed at making decisions common knowledge (Chwe 2001).

as a stoplight, and a Maghribi merchant is expected to hire only an honest Maghribi agent and to reciprocate in sharing information. Social rules also specify the objective function of the team (winning). Social rules define, articulate, and disseminate social positions, objectives, causal relationships, and expected behavior. By providing commonly known cognition, information, and coordination, they delineate causal relationships and expected intertransactional linkages, behavior, and outcomes.

Analyses of the necessary conditions under which learning leads to equilibrium behavior reveals the behavioral implications of commonly known social rules. These analyses indirectly indicate that the properties of learning guided by commonly known rules are very different from those based on individualistic learning. Reaching self-enforcing regularities of behavior requires neither the restrictive conditions of classical game theory nor those of individualistically based learning models. Indeed, reaching regularities of equilibrium behavior rests on intuitive assertions.

Kalai and Lehrer (1993a, 1995) considered learning in a repeated game in which individuals share a cognitive system but each knows only his payoff matrix and discount factor. In other words, the players have the same cognitive understanding of the situation, do not know the relevant parameters of the model, and the objective of each is to maximize his payoff. Observing the outcomes of the game, each player can develop his own subjective evaluation of these parameters and others' strategies. The analysis also assumes that individuals are subjectively rational, in the sense that they start with subjective beliefs about the strategies used by each of their opponents. There is no assumption that each player believes the others are rational. Each individual then uses these beliefs to compute his own optimal strategy.

Analyzing the process of learning reveals that one of the main requirements for convergence on regularities of behavior is a restriction on each player's initial subjective beliefs about other players' strategies. If each player's initial subjective beliefs assign a positive probability to the events that will indeed occur in the play of the game, then eventually learning will lead each player to be able to predict the behavior of the others. Furthermore, these players

will converge in finite time to play a Nash equilibrium of the real game.[8] Subjectively developed beliefs converge on equilibrium beliefs. An initial "grain of truth" regarding others' behavior is thus sufficient for individuals to learn independently how others will play and for convergence on an equilibrium.

The implication regarding the role of social rules in leading to regularities of behavior is clear: an initial "grain of truth" regarding others' behavior provided by social rules is sufficient for individuals to independently learn how others will play and for convergence on a (Nash) regularity of behavior. *Social rules help individuals form beliefs—represented by probabilistic estimates—about the situation and what others will do*.[9] As long as subjectively rational individuals accept the behavior associated with the social rule as possibly correct and respond based on their private knowledge and information, learning will lead to a regularity of behavior (specifically, to a Nash equilibrium).[10] Furthermore, a social rule that correctly informs each individual about how others will actually play is a sufficient condition for a Nash equilibrium, even if the player has neither a complete model nor the ability to make the necessary calculations to find the equilibrium set. If the rule is correct, it must be the case that each player, responding to the rule based on his private information and knowledge, finds it optimal to follow the rule.

---

[8] Specifically, they will learn to play an equilibrium that satisfies the Nash restrictions or those of the epsilon-Nash equilibrium, but such details are not important here. Although the argument in Kalai and Lehrer (1993a) is intuitive, the technical analysis rests on the assumption that individuals use Bayesian updating in response to new information. In fact, people may not update their beliefs based on Bayesian reasoning. If they do not, however, social rules specifying other players' behavior are arguably even more important in leading to regularities of behavior. Indeed, a sufficient condition for a Nash equilibrium is that every individual should have an accurate prediction of what others will do rather than knowledge (or common knowledge) of the rules of the game (Aumann and Brandenburger 1995). The intuition is that a Nash equilibrium is a strategy combination in which each player's strategy is optimal for him, given the strategy of the others. If each player knows what others will play and nevertheless finds it optimal to behave as expected of him, the rule of behavior must satisfy the Nash condition.

[9] See Schotter (1981, p. 52) on the informational role of norms. See Lewis (1969); Sugden (1986; 1989); and Young (1993, 1998) on the informational role of conventions. These analyses focus on individualistic learning rather than the role of institutionalized rules in guiding behavior.

[10] One should not confuse formal rules with institutionalized ones. The formal rules of the road set speed limits, but after watching how fast experienced drivers go, new drivers usually do not adhere to these limits for long. The formal rule helps drivers form beliefs, which they update based on observed behavior.

In situations in which institutions generate behavior, social rules correctly inform each individual how others will behave because of their dual nature as exogenous to each individual whose behavior they influence but endogenous to all of them. They are exogenous to each individual in the sense of being commonly known. But because each individual is playing against these rules, they aggregate private information and knowledge through each player's response to them.[11] In situations in which institutions generate behavior, social rules and the associated beliefs therefore constitute an equilibrium. Each individual, relying on the social rule to enable and guide his behavior—to form beliefs about others' behavior and his best choice of action—finds it best to follow the rule.

In such situations one does not need to know more than this social rule, because *institutionalized rules aggregate private information and knowledge and distribute it in compressed form*. If, for Hayek, institutions constitute a "device for coping with our ignorance" (1976, p. 29), this device manifests itself in institutionalized rules. Institutionalized rules are a useful device because they provide the cognition, information, and coordination required for choosing behavior. They span the domain within which one can make rational decisions. At the same time, institutionalized rules aggregate the knowledge and information of the interacting individuals. In doing so, they direct individuals to play an equilibrium outcome.

This role of Institutionalized rules is well recognized for the particular case of market prices. They aggregate market participants' private information and correspond to an equilibrium outcome. Taking market prices as given, each economic agent responds based on his private information. Hence unless prices already incorporate all of this private information, they cannot be in equilibrium. The response of economic agents will cause the quantity demanded to differ from the quantity supplied, causing prices to change. In equilibrium, prices provide a sufficient statistic for each individual to make an informed, optimal decision. At the center of the argument is the relationship between a public signal and each individual's response to it.

A similar relationship between a public signal—institutionalized rules—and each

[11] "Playing against the rules" means neither playing in accordance with nor playing in violation of the rules. It means that each individual takes the commonly known social rules as exogenous and bases his behavioral choices on the content of these rules as well as his private knowledge, information, and preferences.

individual's response to it, more generally, is at work in situations in which institutions generate behavior. Institutionalized rules provide coordination, and they aggregate and disseminate knowledge and information. The only social rules that can correspond to actual behavior are those in which each individual, basing his decision on his private knowledge and information, finds it optimal to follow the rules. Hence in an institution, institutionalized rules aggregate the private knowledge and information of all agents, providing each with a sufficient statistic to make an informed decision.[12]

Behavior in competitive markets theoretically aggregates information correctly; this is not necessarily the case for institutionalized rules in general. When information is revealed through behavior in social interactions, the information aggregation process depends on the prevailing self-enforcing behavior, which, in turn, depends on the available information. If the players believe that the time discount factor cannot support cooperation in a repeated prisoners' dilemma game, for example, they will not cooperate and therefore never find out that this is not the case. Although one knows his discount factor, others' discount factors are not revealed to him by their behavior.[13]

In situations in which institutions generate behavior people are motivated to acquire the relevant public signal —social rules—just as they are motivated to learn about prices in market situations. In the market each individual is motivated to discover what are the prevailing prices because of their informational value. More generally, individuals interacting in situations in which institutions generate behavior have an incentive to discover the prevailing rules of behavior, because they reflect an equilibrium and hence following them is one's best response. In deciding how to act and when forming beliefs about others' behavior, individuals respond to

---

[12] This is not to argue that rules precede beliefs in the process of institutional emergence; as the discussion in Part III emphasizes, beliefs often precede rules. Similarly, there is not necessarily a process of learning and a convergence, as the issue of choosing which side of the road to use when driving reveals. The argument here is about the system of rules and beliefs that can be institutionalized.

[13] Kuran (1995) emphasized that private information is often distorted in situations in which institutions generate behavior. Individuals are deterred from correctly revealing information about their preferences given the information about others' preferences revealed by this behavior. An individual is motivated to falsify the public representation of his preference as this is the best response to the information revealed by the rules.

socially transmitted rules that they believe come from a reliable source. Doing otherwise is costly, and at times the implications are even irreversible: one may not have many opportunities to find out if individuals are expected to drive on the left or the right or how others will act at an intersection.[14]

The Maghribi traders and the German merchants whose behaviors were examined in Chapters 3 and 4 did not have to solve the mathematical models we now use to study their institutions—nor did they have the information required to do so. Yet each was motivated to learn and was guided by a simple socially transmitted rule of behavior to which he responded based on his private information and knowledge. Game-theoretic analysis is useful in considering this feedback, because it captures the response of each individual to the shared beliefs—created by social rules—about how others will play and restricts the set of these beliefs to be an equilibrium.

We can thus see how institutionalized rules and the beliefs they help form enable, guide, and motivate most individuals to adopt the behavior associated with their social position most of the time.[15] An individual adopts the appropriate behavior because other members of the society condition their behavior on the individual's social position; given the others' expected behavior, an individual's best response is to behave in the way others expect him to do.[16] Socially constructed characteristics— social positions—have behaviorally meaningful implications,

---

[14] This is not to say that institutions do not endogenously change and people do not seek to alter the prevailing equilibrium. I return to this issue in Part III.

[15] The reasons for and the role of deviators is discussed below and in Part III.

[16] In situations in which institutions generate behavior, individuals are seemingly *rule followers*; they follow the rules associated with the social positions they occupy. March and Olsen (1989) argue that peoples' tendency to adopt the behavior associated with their position does not reflect an instrumental logic that asks, "What is my interest in this situation?" Instead, this tendency reflects a "logic of appropriateness," which asks, "Given my role in this situation, what is expected of me?" March and Olsen argue that individuals behave "appropriately" out of a sense of social obligation rather than the promise of reward or the fear of punishment. Such intrinsic motivation is critical, as I discuss later and is easily integrated in the framework developed here. But the mere observation that people seek to find out and then follow the behavior associated with the and ir roles does not reveal the logic behind it.

because equilibrium behavior and expected behavior are conditioned on them.[17] The king's strength comes not from his army but from the beliefs held by each member of the army that everyone else will obey the king's orders and that the best response is also to obey. In situations in which institutions generate behavior, rules of behavior are both prescriptive and descriptive; institutionalization is complete when the behavior associated with the institutionalized rules becomes routine, habitual, and taken for granted.[18]

Whether the private or social propagation of rules will better prevail—and hence be more likely to prevail—depends on the structure of the situation. When this structure is such that an individual who does not know the relevant rules imposes an externality on others, rules are better propagated socially through a dedicated public organization. Because society does not want every new driver to figure out the rules of the road through experimentation, it mandates that a public organization establish the rules of the road and disseminate them. When there is no such externality, rules are more likely to be propagated privately, based on the incentives of individuals to study or transmit them. Among the Maghribis, fathers taught their sons the appropriate rules. When institutionalized rules serve the interest of particular social units—parents, the state, the church, priests, corporate CEOs—they will labor to propagate these rules.

Public propagation of institutionalized rules also takes place because the role of many institutions is deterrence, and actual punishment is socially costly. Furthermore, complex institutions of this sort often rely on the coordinated response of many to a deviation. In situations in which it is the expected reactions of the many that influence one's decision how to act, it is imperative that the understanding of the circumstances under which individuals have to act be common knowledge. It is imperative that many individuals attribute the same meaning to an objective situation or action.

---

[17] As Calvert (1995, p. 59) notes, if "the underlying game does not set apart any individual players as having special opportunities or powers, then such role differentiation can be maintained only as part of an equilibrium."

[18] In sociology, institutionalization is considered the process in which social practices become sufficiently regular and continuous to be described as institutions (Abercrombie et al., 1994, p. 216).

This role of rule was already noted in Chapter 3, which argued that the credibility of the threat in collective punishment among the Maghribis would have been undermined without a merchants' law that defined a common, shared, ex ante understanding of what actions constituted cheating (Greif 1993, p. 542). The study of the Hanseatic League also reflects the importance of a shared understanding of the meaning of various actions. Institutional failures in this case led to organizational changes; as Chapter 4 showed, the embargo of 1360 ushered in a long period in which no conflict occurred between Bruges and the Hansa. This outcome was due partly to a change in the underlying cognitive foundation. The merchants' privileges were written "in much detail as to prevent any one-sided interpretations" (Dollinger 1970, p. 66). Fearing the responses of many merchants, agents and rulers did not cheat or abuse property rights. The shared meaning of various actions was therefore crucial to making this collective response credible.

Rules specifying the meanings of various actions (i.e., whether a transgression has occurred or not) are general features of situations in which the threat of collective responses influence actions. Social pacts, customary laws, constitutions, and traditions are among the manifestations of rules that, by creating common knowledge, lend credibility to such threats.[19]

In situations in which institutions generate behavior, rules disseminate a shared cognitive system (including the specification of social positions and states of nature), which is needed to specify and transmit behavioral rules whose information content and coordination functions help individuals form beliefs about what others will do and hence choose their behavior. Each individual, seeking guidance for behavior at the social level, is motivated to learn them. Each individual responds to these rules based on his private knowledge and information, leading to the aggregation of knowledge and information. In situations in which institutions generate behavior, each individual finds it best to follow these rules, and because each individual behaves as expected, no information is generated to cause individuals to change their behavior. Regularities

[19] Shared meaning and the collective responses that such meaning renders possible also provide the institutional foundation of the state. The Magna Carta offers an example of how institutionalized rules provide the institutional foundations of the state by creating the shared meaning required for beliefs that political agents will collectively respond to a transgression of their rights by a ruler. For analyses in this spirit of modern political systems, see Hardin (1989); Prezworski (1991); and Weingast (1995, 1997), among others.

of behavior prevail, and players hold accurate beliefs about others' behavior, even though they lack a complete model or the ability to deduce other's behavior.

5.2 Game Theory and Modeling Endogenous Institutions

We can now see why and to what extent game theory is a useful tool for studying behavior generated by institutions. The game-theoretic assumption that the rules of the game are common knowledge captures the cognitive and informational roles of social rules. The focus on strategies —plans of behavior—that are common knowledge captures the coordinative role of social rules. The game-theoretic analysis captures the idea that in situations in which institutions generate behavior, social rules provide players with a common cognitive model, information, and coordination that enables each individual to form beliefs about others' behavior. It restricts the rules that can prevail to those that are self-enforcing, where each individual, expecting that everyone else will follow the rules, finds it optimal to follow them as well.

The games used to study the Maghribis' coalition and the merchant guild embodied cognitive aspects, such as traders, merchants, agents, cheating, rulers, territory, money, penalties, cities. The models also assumed that the players had the required knowledge to condition their actions on these cognitive aspects—that the Maghribi traders knew how to recognize each other and shared a common understanding of what behavior constituted cheating, for example. The analysis of the merchant guilds assumed that the merchants were informed about aspects of the situation, such as the territorial area of a ruler, who the representative of the *Kontor* was, who a merchant from a particular town was, and so forth. Simple rules of behavior enabled and guided the behavior of merchants, agents, and rulers.

At the same time, game-theoretic analysis restricts the set of admissible social rules that can be common knowledge and correspond to behavior exactly by demanding that these rules aggregate private knowledge and information. The game-theoretic analysis restricts the set of behavioral beliefs that can be common knowledge, correspond to behavior, and are not refuted by it. The analysis achieves that by considering possible equilibria. When beliefs that the interacting individuals hold are commonly known and each player plays his best response to them (and is rational in this limited sense), the set of beliefs is restricted to those associated with an

equilibrium behavior. In other words, admissible behavioral beliefs and the corresponding coordinative rules are those that are self-enforcing (Greif 1994a, p. 915).[20] Nash equilibrium analysis restricts beliefs about behavior on the equilibrium path, that is, in circumstances that can transpire with positive probability given the expected behavior. Equilibrium refinement concepts, such as subgame perfection, restrict beliefs about behavior off the equilibrium path, that is, in circumstances that will not transpire given the expected behavior. Using the subgame perfection equilibrium concept has the intuitive appeal of restricting expected promises and threats off the equilibrium path to those that are credible. The Nash restriction (on behavior on the equilibrium path) also limits the set of admissible behavioral beliefs, and hence institutionalized rules, to those that are reproduced, not refuted, by the implied behavior. Nash equilibrium requires that individuals correctly anticipate one another's behavior, and hence they do not encounter behavior that refutes their expectations.

Game-theoretic equilibrium analysis restricts the set of institutionalized behavioral rules and beliefs (including beliefs about intertransactional linkages) that guide and motivate behavior given the cognitive content of the institutionalized rules. At the same time, it restricts the set of admissible internalized beliefs — the cognitive content of institutionalized rules — to those that are reproduced, not refuted, by the resulting behavior. Applying the Nash restriction does not directly limit the cognitive structure imputed in the rules of the game. Indeed, there is no theory that deductively restricts the admissible cognitive structures in a given situation.[21] Yet because

---

[20] Formally, in a complete-information, extensive-form game, denote by *P* a path of play, and define S(P) to be the set of all strategy combinations for which the path of play is P. Denote the beliefs of player i by $B_i(S(P))$, defined as a probability distribution over S(P). Note that the possible probability distributions differ only in the weight they place on different behaviors off the path of play. Concentrating on this probability distribution thus captures the notion that the player has to hold the beliefs generated by the observation that a particular path of play is followed. Denote by $B(s^*)$ the shared beliefs that strategy combination $s^* \in S(P)$ will be played. That is, $B_i(S(P)) = \{\text{Prob}(s^*(P) = 1\} \; \forall$ i for $s^* \in$ S(P). When $U_i(s_i^*, B(s^*)) \geq U_i(s_i, B(s^*)) \; \forall$ i and $\forall \; s_i \in S_i$ (i.e., following this strategy is the best response given the beliefs), then s^* is a Nash equilibrium. Hence $s^*(P)$ is an equilibrium, and the associated beliefs are self-enforcing.

[21] Reviewing the vast literature in cognitive science, Mantzavinos (2001) argues that we are not likely to develop such a theory. Kaneko and Matsui (1999) and Aoki (2001) developed inductive game theory, which explores whether purely individualistic behavior can generate regularities of behavior when each of the interactive individuals inductively develops his own subjective understanding of the

such an analysis exposes the relationships between the rules of the game and possible outcomes, we can restrict admissible models to those in which the implied behavior reproduces—does not refute—the cognitive models imputed in the game.[22]

The logic of reproduction of the cognitive model—its confirmation by the observed outcomes—must have been on the mind of the prophet Elijah when he challenged the pagan priests of Baal to call upon their idol to light a fire on his altar on Mount Carmel. Their failure to demonstrate their idol's ability in this way cost them their lives and convinced the Israelites to return to worshiping God. The repeated failure of the merchant guilds to protect the property rights of the German merchants in Bruges refuted the merchants' beliefs that rights would be respected. Institutional change followed.

A game-theoretic analysis therefore evaluates whether, given our perception of the objective structure of the situation, the assertion that a particular institution—consisting of particular rules and beliefs— is logically consistent. The analysis restricts institutionalized rules by limiting the set of admissible beliefs and behavior to those that are self-enforcing and reproducing. (For simplicity I henceforth refer to such institutions as self-enforcing and denote reproduction separately only when the distinction between the two concepts is important.)

Technically, presenting a situation as a game entails specifying the rules of the game, the relevant actors, their actions, the information available to each when choosing behavior, and the outcomes associated with various behavioral choices. The discussion presented here, however, highlights that conceptually, when we present a situation as a game, we are making a statement about our own understanding of the objective features of the situation, our perception of the relevant intertransactional linkages, and the cognitive and informational content of the prevailing institutionalized rules. A model constitutes a statement about the players' understanding of the situation (Rubinstein 1991).

When interpreting the analysis, therefore, we have to keep in mind that while we study games, real-world actors do not play against the (commonly known) rules of the game but against

---

situation.

[22] As is well known, individuals tend to interpret evidence in a way that confirms their prior beliefs. Part III considers the implications of this tendency for institutional change.

commonly known institutionalized rules.[23] The Maghribi traders' coalition was studied as if each individual played against the rules of the game. The analysis indicates that there could have been a cognitive model of the situation consistent with our understanding of it and beliefs about the various unobservable features of the situation (e.g., time preferences and outside options) that could have rendered self-enforcing the beliefs in the rules of intragroup hiring, honesty, and punishment. Clearly, each Maghribi trader did not solve this game-theoretic model, directly observe the factors that were important to others' decisions, or necessarily understand the nature of the institution as an equilibrium outcome. But the analysis substantiated that each trader could have found it optimal to adhere to the associated behavioral rules while responding to the social rule.

Such analyses can be used to capture, when appropriate, the direct and indirect influences on behavior of the actors' internalized belief system regarding the natural and supernatural worlds around them. These internalized beliefs influence the perceived utility of taking a particular action and thereby directly influencing it. Notice that these beliefs may be unverifiable on the path of play. If enough members of a society internalize the belief that God will send a cheater to hell, they may behave honestly. The Aztecs internalized the belief that the world would end if human blood was not shed in the evening. The belief could not be refuted by observable outcomes, because it motivated the Aztecs to shed blood every evening. Outcomes that could have refuted the beliefs were off the path of play and the existence of alternative possible institutions was not revealed.

Internalized beliefs indirectly influence institutionalized behavior by changing the set of self-enforcing behavioral beliefs. If the internalized belief that God sends blasphemers to hell is an institutional element, a borrower can credibly commit to pay his debt by taking an oath to be honest, because breaking the oath would show contempt for God and entail divine punishment. Of course, there can be uncertainty over who internalized such a belief. Such uncertainty is captured in incomplete information models, which reveal that even if the actual number of true believers in the population is small or even zero, it can nevertheless have a large impact on

[23] The game-theoretic implications of this distinction are not yet well developed.

behavior, because nonbelievers find it beneficial to pretend to be believers (see Kreps et al. 1982; Appendix C, and Kuran 1995).

Although institutions generate regularities of behavior, there are usually some individuals who, for idiosyncratic reasons, will not follow the behavior expected of people in their positions. The implied responses to such deviations are important in reproducing institutionalized rules and beliefs regarding behavior off the path of play. Game theory restricts the analysis of this deviation-as-confirmation mechanism in two ways. First, this mechanism operates only if the threats that follow deviations are credible. Behavior and expected behavior have to correspond to a subgame perfect equilibrium that restricts the threat of behavior off the equilibrium path to be credible. Second, game-theoretic learning models explicitly incorporate how individuals update their beliefs about others' behavior in the specification of the game, thereby enabling a study of the limits of the deviation-as-conformation mechanism.

Ironically, the more effective an institution is in preventing deviations, the more individuals are likely to maintain that different rules of behavior will prevail off the equilibrium path. More generally, "semi-institutionalized" situations are those in which there is no uniformity of expectations regarding actions that will be taken off the equilibrium path. On-the-equilibrium-path behavior (where there is uniformity of beliefs) is still self-enforcing and reproducing, and each individuals' best response is to follow the behavior expected of him.[24]

5.3 Institutional Ramifications of Social and Normative Behavior

The discussion so far has ignored the social and normative foundation of behavior.[25] Everything

[24] Subjective game theory (Kalai and Lehrer 1993b, 1995) and self-confirming equilibria (Fudenberg and Levine 1993, 2003) provide an appropriate analytical framework in this case. Roughly speaking, in equilibrium an individual can hold any beliefs about the behavior of others that is not contradicted by the observed implications of the actual behavior and still generates the equilibrium path behavior. For an empirical analysis of such a semi-institutionalized situation, see De Figueiredo, Rakove, and Weingast (2001).

[25] Sociologists have explored this foundation (for reviews, see Wrong 1999 and Scott 1995). Its importance has also been stressed by many prominent economists, including Becker (1974); Arrow (1981); Hirshleifer (1985); Akerlof (1986); Lal (1998); North (1990); Platteau (1994); Samuelson (1993); and Sen (1993). Evolutionary models of the origin of social and normative propensities can be traced back to Wilson (1975). For recent contributions, see Güth and Yaari (1992); Güth (1992); Bowles

else being equal, people seek to act in a manner that generates positive social responses by the people they know, elevates their social status and esteem in the broader society, provides them with identity, and is consistent with their (internalized) norms.

In modern sociology the argument over the behavioral importance of social exchange, beliefs in others' social responses, or losses of esteem following a particular action is associated with Homans (1961), Wrong 1999 [1961], and Granovetter (1985). Another line of research, associated with Talcott Parsons (1951), emphasizes the importance of norms in motivating behavior by influencing the intrinsic utility from it.[26] Internalization of norms, or the incorporation of behavioral standards into one's superego, essentially means the development of an internal system of sanctions, one that supports the same behavior as the external system.[27] In this theory, "values and norms were regarded as the basis of a stable social order" (Scott 1995, p. 40).[28]

Recent work in experimental game theory has convinced even skeptical economists of the importance of the social and normative foundations of behavior. Some individuals do act altruistically - that is, they are willing to decrease their own material welfare if it increases that of others - (Andreoni and Miller 2002; Charness and Grosskopf 2001; Kritkos and Bolle 1999). Knowing another person—even just by sight—alters how altruistic one is willing to be toward that person.[29] Some people exhibit inequality aversion, expressed as concern about the equality of

---

and Gintis (1998); Huck and Oechssler (1999); Bester and Güth (1998); Kockesen, Ok, and Sethi (2000a, 2000b); Ely and Yilankaya (1997); Dekel, Fudenberg, and Levine (1999) and the reviews and contributions in Field (2002) and Gintis (2000). Some evolutionary models, such as Kandori (2003), cast doubt on the long-run sustainability of normative behavior, pointing out that they are likely to be eroded.

[26] Psychologists define an intrinsically motivated act as one that is taken despite the lack of any reward from doing so except for the value of the action itself (see the review in Frey 1997, pp. 13–14).

[27] On norms and their transmission, see K. Davis (1949); Cavalli-Sforza and Feldman (1981); Bandura (1971); Witt (1986); Shapiro (1983); and Elster (1989a, 1989b).

[28] A finer sociological distinction is that between values specifying the preferred or the desirable (e.g., winning the game) and norms specifying the legitimate means of achieving these goals (e.g., winning by playing fair). To simplify the discussion, I use the term *norm* to include both.

[29] Consider the dictator game experiment in Bohnet and Frey (1999), in which the "dictator" can impose any division of ten dollars between himself and another player. Only 25 percent of dictators

the payoffs between themselves and others.[30] Many individuals reciprocate the behavior of others, even if doing so reduces their material well-being. They respond to "fair" behavior, for example, with actions that raise others' material payoffs.

Such social and normative behavior is *situationally contingent*: whether a particular action insults others, how status is acquired, who is deserving of altruism, and what constitutes fair behavior depend on the time and place. As sociologists and anthropologists have long argued, a wide range of behavior is socially and normatively sanctioned. Findings in social physiology (see, e.g., Ross and Nisbett 1991) lend support to this observation.[31] Game-theoretic experiments conducted by E. Hoffman, McCabe, Shachat, and Smith (1994), Henrich et al. (2001, 2004) and Roth et al. (1991) among others reach the same conclusions.[32]

The social and normative foundations of behavior can have institutional ramifications. "Institutions are something beyond us and something in ourselves," wrote Durkheim (1953, p. 129). They are "something in ourselves" when beliefs associated with social responses and expected normative behavior generate regularities of behavior. Studying the institutions within ourselves amounts to studying particular intertransactional linkages. Considering the institutional

---

divided the money equally when the game was played anonymously, but 71 percent did so when the two players were identified to each other. See also Dawes and Thaler (1988); E. Hoffman, McCabe, Shachat, and Smith (1994); E. Hoffman, McCabe, and Smith (1996, 1996a, 1996b); and Ostrom (1998).

[30] Fehr and Schmidt (1999) survey relevant experiments; see also Loewsenstein, Bazerman, and Thomson (1989) and Bolton and Ockenfels (2000). Some individuals are willing to make inequality-increasing sacrifices when they are efficient and inexpensive.

[31] For theoretical support, see Andreoni and Miller (2002). They note the failure to find a general model of social preferences and conclude that "many things other than the final allocation of money are likely to matter to subjects. Theories may need to include some variables from the game and the context in which the game is played if we are to understand the subtle influence on moral behavior like altruism" (p. 20). The axiomatic approach for social preferences led to similar conclusions (Segal and Sobel 2000).

[32] Platteau and Hayami (1998) and Platteau (2000) have argued that environmental factors influence norms. Different norms manifest themselves even in current laws. For empirical evidence from the United States, see Young and Burke (2001). Distinct notions of who is responsible for acting altruistically toward whom are reflected in social welfare policy. Until very recently, the Japanese Civic Code Article 877 specified that family members within three lineal generations had an obligation to pay for the living costs of a disabled family member. This is not the case in the United States, where family members have no such legal responsibility.

ramifications of social exchange amounts to examining the linkage between social and economic transactions; studying norms allows consideration of the "transaction" between an individual's superego and his ego or id.

A way to study such social and normative intertransactional linkages analytically using the game-theoretic framework is to take norms and beliefs associated with social exchange as given and integrate norms and social sensitivities in the specifications of actions and payoffs. Such games allow one, for example, to take a "social" action, such as displaying spite, and to specify the players' preference to be conditional on such actions. The behavioral beliefs and behavior that can prevail as an equilibrium outcome in this extended game are then studied; self-enforcing behavioral rules and beliefs will reflect the actual or perceived social responses of others' reactions to various actions and the psychological cost of acting in ways that are not consistent with one's internalized norms.[33]

We can go farther and use game theory to study the simultaneous determination of behavior and its social and normative foundations. The situational contingency of social and normative considerations implies that people seek social and normative guidance about what is socially acceptable and normatively appropriate. They find this guidance at the society level in the form of social rules that define the means for gaining status, the reasons to resent others, the behavior that is normatively sanctioned, and the normative frame to use in particular situations.[34] Which of these commonly known beliefs about social responses and which norms that motivate behavior can be self-enforcing? What factors influence whether socially appropriate and normative behavior is a cultural phenomenon that does not correspond to behavior or institutionalized rules that do?

Game theory is flexible enough to allow players' preferences to be specified in a way that

---

[33] There is much related economic research (mainly theoretical and focusing on contractual and organizations issues), reviewed in Fehr and Schmidt (1999). Cole, Mailath, and Postlewaite (1992) analyzes the growth implication of how societies bestow social status upon their members. See also Fershtman and Weiss (1993) and Benabou (1994). The difference in normative dispositions among individuals (indicated by experimental game theory) can be incorporated in the analysis using incomplete information games. See Kreps, et.al. (1982).

[34] A framing effect (Tversky and Kahneman 1981) is a change of preferences between options as a result of a change in the formulation of the issue or problem.

captures their sensitivity to others' social responses and the dependency of their norms on the extent to which others adhere to them. At the same time, this specification can and should capture the material costs that such behavior can entail. It thus allows us to model the simultaneous determination of behavior and its social and normative foundations through feedback between each individual's choice of behavior and aggregate behavior. Using game theory, we can identify the factors that influence socially appropriate and normative behavior by considering which social and normative rules of behavior can be common knowledge, which can correspond to an equilibrium behavior, because each individual responds to them while taking into account the material cost of following them whenever appropriate.

As an example, consider the analysis of Höllander (1990), who integrates social exchange theory in examining voluntary cooperation in the provision of public goods.[35] He assumes that individuals respond to emotionally prompted social approval and that the desire to gain social approval influences economic behavior. When choosing behavior, each individual considers the economic cost of contributing a particular amount to the public good as well as the social approval and disapproval associated with doing so. The social approval or disapproval that a particular action implies is determined by the actions other individuals have taken. Specifically, the social approval or disapproval is proportional to the difference between one's contribution to the public good and the others' average contribution. In the game-theoretic equilibrium, an individual's behavior is influenced by self-enforcing behavioral beliefs about how much others will contribute and the implied trade-off between the desire for social approval and the cost of providing a public good.[36] Annex 7.1 in Chapter 7 provides an example of a game-theoretic

[35] His analysis therefore examines the implication of linking an economic transaction with a social one. As noted below, institutions also influence whether a contribution to a particular public good confers esteem or not.

[36] Psychological game theory (Geanakoplos, Pearce, and Stacchetti 1989) studies endogenous psychological motivations, such as anger and pride, by assuming that utility functions are belief dependent. "The players' payoffs depend not only on what everybody does but also on what everybody thinks" (p. 61). Equilibrium beliefs correspond to reality and deviate from expected equilibrium behavior can trigger an emotional response. One's behavior is influenced by self-enforcing beliefs about others' emotional responses, and these beliefs are reproduced by the implied behavior. Applicability is limited by the problem of multiple equilibria (see Rabin 1993; Fehr and Schmidt 1999; and Charness and Rabin 2002, for a game-theoretic evolutionary approach to norms, see Frank 1987).

analysis of social exchange.[37]

As this discussion illustrates, a useful feature of game theory is that it allows us to study all intertransactional linkages—economic, coercive, social, and normative—simultaneously using the same analytical framework. Such an integrative framework responds to the concerns of the eminent sociologist Dennis H. Wrong (1999), who argues that taking the social and normative foundations of behavior as exogenous is too simplistic. We cannot, according to Wrong, "dispense with the venerable notion of material 'interests' and invariably replace it with the blander, more integrative 'social values'" (p. 43). Recognizing the importance of normative considerations "does not mean that [they] have been completely molded by the particular norms and values of their culture" (pp. 45–6). What is needed is an integrative framework that captures the fact that various factors—social, normative, and materialistic—can simultaneously influence behavior. Game theory provides such a framework, one in which social exchange, norms, and materialistic considerations (regarding money, power, and other materialistic rewards and sanctions) can easily be integrated.

In a game-theoretic analysis, payoffs can be conditioned on the actions taken to reach a particular outcome and the players' beliefs regarding appropriate and emotional responses. This attribute of the game-theoretic framework renders it useful for studying the normative and social foundations of institutionalized behavior. These considerations can be incorporated in the rules of the game to examine their impact on behavior and behavioral beliefs. They can also be derived endogenously as equilibrium outcomes.

5.4 Legitimacy and the Origin of Institutions

Because institutions are equilibrium phenomena, it is conceptually sound and analytically useful to discuss them without examining their origins. Whether an institution evolved spontaneously or was established intentionally, whether it reflects individualistic learning, evolutionary pressure, or social design, its equilibrium nature is the same. I touch on some aspects of institutional origin in Chapter 7. Here the discussion focuses on the way rules are mapped into beliefs, as it is such

[37] Aoki (2001) provides a game-theoretic analysis of social exchange in premodern Japan.

mapping that differentiates social from institutionalized rules. Social rules are commonly known, whereas institutionalized rules are social rules that are commonly believed to be followed.

For an institution to be established by decree, it is necessarily that a sufficiently large number of those who are supposed to follow the rule believe that others will follow it. Each individual must believe in the cognitive content and coordinative impact of the rules and/or believe that its declaration will affect social exchange and norms. If individuals do not hold these beliefs, they will not follow the rules, even if the rules correspond to an equilibrium (i.e., specify self-enforcing beliefs, norms, and behavior). Unless a rule leads to beliefs that it will be followed, the behavior it prescribes will not be followed. The legitimacy of those who issue rules is therefore central to institutionalization. Indeed, in the absence of individuals or organizations with such cultural authority, institutions would never emerge by decree. All institutions would emerge from individualistic learning processes, which economists model well (see, e.g., Chamley 2004). Arguably, however, complete inability to coordinate by decree is not optimal.[38]

In most societies some social units have the legitimacy required to alter institutions. One universal source of legitimacy is the observation that rules issued by the social unit in the past have been followed. The individuals and organizations with legitimacy, however, differ across societies and situations reflecting initial conditions including organizational heritage and internalized beliefs. Once established, a social norm conferring legitimacy constitutes an equilibrium: if it is expected that a new legitimately issued (equilibrium) coordinating rule will be followed, it will be. The more such new rules are followed, the more they will confirm the legitimacy of those who issued them. Because different legitimate authorities are likely to have different objectives and because societies differ in terms of their legitimate authorities, institutional development is likely to vary across societies.

Legitimacy is therefore central to institutional development. But contemporary students of

[38] I am not familiar with general analyses exploring the trade-off between inability to coordinate and the influence of the coordinator on the efficiency of the resulting institution. Hayek (1979) stresses the importance of rules issued by decree.

institutions in economics, political science, and economic sociology have little to say about it.[39] Accordingly, I note here only that the late medieval period was crucial in Europe in terms of the development of legitimacy norms. During this period, rulers were well aware of the value of legitimacy in facilitating their rule and preventing challenges. Legitimacy is at the heart of the Bayeux Tapestry (1092), for example, which depicts how the Normans, led by William, conquered England in 1066. The tapestry was ordered by Odo, William's half brother and the bishop of Bayeux. Its opening scene shows Edward the Confessor, the last Saxon king, bestowing the kingdom upon William, thereby establishing William's legitimacy. After conquering Sicily and southern Italy, other Normans sought legitimacy by giving the area to the pope and ruling as his vassals.

These examples reflect the struggle between the secular and the religious regarding the source of legitimacy of rulers and rules in medieval Europe. During the late medieval period, the church was in the process of losing its bid to become the ultimate source of legitimate rules governing the practical aspects of the polity, society, and economy, either by nominating rulers or issuing rules. The beliefs in the appropriateness of man-made customary law embedded in Roman law and customary German law, which suited the interests of traditional secular leaders who therefore cultivated it, played an important role in this process. The failure of the church to prevent its members from strategically using their canonical position for their material benefit may have been instrumental as well by undermining the moral foundations of the church legitimacy (Ekelund et al. 1996).

During the late medieval period, legitimacy norms increasingly rested with the state in Europe. Rules were legitimate if issued by rulers with the hereditary right to the throne, conferred through a participatory process of selection, or issued through a participatory process of rule making. The Magna Carta, the elected monarchy in Germany, the Swiss confederacy, the Italian city-states, and the French Estates-General are among the many manifestations of this process,

---

[39] The term *legitimacy* has only six index entries and receives very little coverage in the *New Handbook of Political Science* (Goodin and Klingemann 1996). The *Handbook of Economic Sociology* includes no index entries for *legitimacy* (Smesler and Swedberg 1994), although Weber (1947) studied it. Levi (1988) indirectly touches on the issue by discussing "consent" for taxation. The discussion here builds on Greif (2002).

which reached its zenith with the modern democratic state.[40] In the West today, the state and participatory professional associations are the main sources of legitimate rules (DiMaggio and Powell 1991b; Scott 1995).

In the Muslim world the opposite process was taking place regarding legitimacy norms. Early on rulers were legitimated by virtue of being closest to the prophet. Later a ruler's legitimacy increasingly became faith-based, conditional on the ruler respecting, advancing, and promoting Islam. Failure to do so legitimized the use of force to overturn the ruler. As one of the most esteemed Muslim jurists, al-Mawardi (d. 1058), declared, one should not obey even a caliph if his orders contradict the teachings of Islam. The state, however, had only limited legitimacy as an interpreter of the *Shari'ah*, the Islamic code of law. By the late medieval period, the religious scholars, had already became the legitimate interpreters of the *Shari'a.* Even a caliph had no such legitimacy. Ever since, Islamic rulers have attempted, with various degrees of success, to create a state-controlled *ulama.* Rulers were particularly successful to influence regarding matters that concerned them the most such as taxation and fiscal policy. (E.g., Sonn 1990; B. Lewis 1991; Abou El Fadl 2001; Crone 2004; Kuran 2005).

Yet, the need to circumvent, evade, or confront this source of legitimacy influenced institutional development in the Islamic world. Indeed, even when monarchies, republics, and dictatorships were established in the Arab Middle East after the demise of colonialism, the traditional sources of legitimacy still held sway. Even these relatively secular polities found necessary to signal their adherence to the *Shari'a*. The constitution of the Egyptian monarchy, established in 1922, for example, declared the *Shari'a* as the source of law. The 1971 constitution of the Arab Republic of Egypt defines Egypt as a socialist democratic state but declares that the principal source of legislation is the *Shari'ah*.

---

[40] The participatory nature of these bodies may directly contribute to their legitimacy. Ostrom (1998, p. 7) surveys experimental evidence indicating that when people can communicate and agree on rules of behavior, they behave in the way that is agreed upon, even if it is not in their material best interest to do so. Stewart (1992) notes that legitimacy confers normative value. In his comparative study of rules regulating the donation and selling of human blood, he notes that where legal rules prohibit the sale of human blood for medical purposes but encourage donations, stronger norms exist against selling blood.

5.5 Concluding Comments

This chapter uses insights from classical and learning game theory to better understand the roles and interrelationships between various institutional elements and the merit, manner, and limitations of the game-theoretic framework for studying endogenous institutions. These insights highlight the importance of institutionalized rules that enable and guide behavior by helping individuals form beliefs about the world around them, about what others will do, and about what is morally appropriate. They create shared cognition, provide information, enable coordination, and indicate morally appropriate and socially acceptable behavior. Individuals seek guidance regarding the situation and how to behave in it; social rules provide this guidance. Social psychologists have convincingly argued that evolution has fine-tuned the human brain's capacity to take actions in situations in which individuals are guided by social rules (Tooby and Cosmides 1992).

At the same time, because retrospective individuals respond to social rules based on their private knowledge and information, institutionalized rules—social rules corresponding to regularities of behavior—aggregate private information and knowledge. The only social rules that can be institutionalized are ones that, if they are expected to be followed and to specify the morally appropriate course of action, are indeed followed and are not refuted by the outcomes these rules, beliefs, and norms generate. An institution can therefore be defined as comprising cognitive, coordinative, informational, and normative social elements that jointly generate a regularity of (social) behavior by enabling, guiding, and motivating it.

Game theory is a useful analytical tool in situations in which institutionalized rules prevail, because such rules correspond to the game-theoretic assumption regarding common knowledge. The analysis then restricts the set of admissible social rules that correspond to behavior to those that are self-enforcing: every individual, believing that others will follow the rules, finds it best to do likewise, given his private knowledge and information. The set of admissible institutionalized rules is thereby restricted. Indeed, self-enforceability in the Nash sense also implies that behavior reproduces—does not refute—the beliefs and does not subvert the norms that motivated it. Social rules that are self-enforcing are the only ones that can be institutionalized. The ability to restrict the set of admissible beliefs is thus central to the way

game theory proves useful for institutional analysis.

The argument developed in this chapter requires further development in many ways. The analytical tools for deductively restricting internalized beliefs (mental models) and norms are limited. Also undeveloped is the argument that people play against (the cognitive and informational content of) rules rather than against the rules of the game. More broadly, as Simon (e.g., 1955) argued, the substantiative implications of limited cognition and rationality are yet to be fully worked out. Further development may benefit from linking strategic behavior with that of the individual seeking to "satisfice" rather than optimize.[41]

Social psychologists argue that behavior is also psychologically motivated, because acting in a way that is at odds with one's conception of one's self is psychologically costly. Moreover, individuals tend to develop identities that correspond to what others expect of them. An honest person develops an identity that renders cheating more difficult; an individual who is expected to be an entrepreneur derives satisfaction from being one. The behavior generated by institutions and the beliefs motivating it therefore lead to corresponding identities and psychological motivation to follow this behavior. Further exploration of the interrelationships between external and intrinsic motivations along these lines seems promising.

Yet, even without these developments, it is imperative to understand the basic interplay between rules, beliefs, norms, and behavior in situations in which institutions generate behavior. People seek cognitive models and information on which to base their behavioral decisions; they seek a means to coordinate their behavior and search for guidance on what is socially acceptable and normatively appropriate. Socially distributed and disseminated commonly known rules provide these microfoundations, enabling an individual to gain cognitive understanding of the situation and information, determine the morally appropriate and socially accepted behavior, and form beliefs about others' behavior. Each individual, however, responds to the commonly known

---

[41] Simon (1955) uses the word *satisficing* (a blend of *sufficing* and *satisfying*) to characterize algorithms that deal with conditions of limited time, knowledge, or computational capabilities. He postulates that an individual will choose the first alternative that satisfies his aspiration level rather than calculate the probabilities of all possible outcomes and choose the best alternative. For a recent survey, see Conlisk (1996). Gilboa and Schmeidler (2001) present an analytical framework for studying *satisficing* behavior.

rules based on his private information and knowledge, implying that institutionalized rules aggregate and disseminate such information and knowledge. In situations in which institutions generate social rules, beliefs, norms, and behavior constitute a system in equilibrium. The game-theoretic framework is a useful tool for institutional analysis because it captures this interplay between rules, beliefs, norms, and behavior, thereby enabling us to restrict the admissible set of institutions.