Part III Institutional Dynamics as a Historical Process

How does an institution persist in a changing environment? How do exogenous changes and the processes that an institution unleashes lead to the institution's demise? How do past institutions—perhaps even institutions that are no longer effective in influencing behavior—affect the direction of institutional change? Why do societies evolve along distinct institutional trajectories, and why is it so difficult to alter institutional dynamics to induce better outcomes?

These questions have long bedeviled institutional analysis in economics, political science, and sociology. Addressing them requires a framework that can accommodate both stability and change—a framework that can account for an institution's persistence and stability in a changing environment on the one hand and endogenous institutional changes and the limit on institutional persistence on the other. The framework must also facilitate studying why, how, and to what extent past institutions influence subsequent ones.

Since the 1970s economists have developed two perspectives—the intentionally created perspective and the evolutionary perspective—to study institutional dynamics. The intentionally created perspective postulates that institutions are intentionally established by forward-looking individuals to serve various functions. Institutional dynamics are best studied as reflecting responses to the functions the institutions serve (e.g., North 1981; O. Williamson 1985).[1] Political economy models were found to be particularly useful in studying processes through which institutions are established and changed. Economic institutions (which in political economy models are defined as formal rules regulating economic activities) are outcomes of political processes; they therefore change following exogenous changes in the decision-making

[1] Institutions reduce uncertainty (Sugden 1989; North 1990), influence distribution (Olson 1982; Knight 1992), maximize groups' welfare (Ellickson 1991), and minimize transaction costs (Williamson 1985). Such functionalist analysis is persuasive only when it is possible to delineate the mechanism linking the origin of the institution and its presumed effect (Stinchcombe 1968, pp. 87–93; Elster 1983; and Fligstein 1990).

process or the political actors' interests (for surveys of this literature, see Weingast 1996 and Peters 1996).[2]

In the intentionally created perspective, the past per se does not constrain institutional changes that forward-looking agents would initiate. The cost of change, rather than the shackles of history, limits institutional adjustments. Institutions fail to adjust in response to exogenous changes, due mainly to sunk costs, coordination costs, and network externalities (North 1990); the costs of overcoming the objections of those who benefit from the existing institutions (Olson 1982); and the difficulties associated with co-opting potential losers (Fernandez and Rodrik 1991, Kantor 1998).

To further the limited ability of this perspective to account for the lack of institutional change, scholars have invoked the stickiness of informal institutions. The argument is that informal institutions—defined mainly as customary rules of behavior, social relationships, or norms—cannot be changed by fiat, and this limits the effectiveness of changing formal rules (North 1990, 1991; Mantzavinos 2001; Aoki 2001).[3] This position is unsatisfactory, however, because, as O.Williamson (2000) notes, it accounts for institutional change by using one analytical framework whereas it accounts for the lack of change by invoking forces outside that analytical framework. Invoking the constraints imposed by informal institutions on the process of institutional change is appropriate only when the forces contributing to the persistence of these informal institutions are explicitly integrated into the analysis (as in Greif 1994a and Ensminger 1997).

Evolutionary Institutionalism, which is rooted in Old Institutionalism and Austrian Economics (Menger 1871 [1976]; Hayek 1937), presents another approach for studying intertemporal relationships among institutions. It usually defines institutions as patterns of behavior reflecting the unintentional consequence of interactions among individuals with limited

---

[2] Transaction-cost economics (following O.Williamson's seminal 1985 contribution) also considers institutions (which are identified with contracts and organizations forms) to be determined by their function.

[3] The general conscription introduced by Britain during World War I exempted the Irish, whose anti-English norms implied that the cost of enforcing conscription and effective military service would be too high (Levi 1997).

rationality. It rejects the forward-looking and functionality premises of the intentionally created perspective.

In formal models capturing this idea, mutation, selection, and inertia link the behavior of limitedly rational individuals with institutions.[4] Each individual is endowed with a trait that dictates his behavior. The relative payoff to a trait depends on the environment and the population distribution of behavioral traits. Selection and the exogenous introduction of new traits—mutation—alter the population distribution of behavioral traits. Over time, more successful traits increase their proportion in the population.

While mutation and selection influence the direction of change in the distribution of traits, inertia determines its rate. The proportion of more successful traits increases only over time. It takes time for selection, operating through imitation or a higher reproduction rate, to transpire. The analysis considers the conditions under which a stable distribution of traits—an equilibrium—is reached.

Stability and change can be studied within the same analytical framework in such evolutionary models, but their microfoundations are restrictive, as noted in Chapter 1. The framework postulates that individuals are not forward-looking; at best they are retrospective. The social level is ignored, as individuals are assumed to be unable to coordinate, communicate, or collectively alter the environment within which they interact. Processes of mutation that drive institutional change are taken as exogenous, while inertia, which determines the rate of change, is assumed rather than derived endogenously.

Chapters 6 through 9 outline another perspective on institutional dynamics. The *institutional dynamics as a historical process* perspective makes explicit the forces contributing to institutional persistence in a changing environment. It exposes when and why institutions endogenously change and how past institutions influence subsequent ones. This historical-process perspective bridges the gap between the Old Institutionalism evolutionary perspective and the New Institutionalism intentionally created perspective. It incorporates the Old

[4] See, for example, R.Nelson and Winter (1982); Frank, (1987); Sugden, (1989); Young (1993, 1998); G.Hodgson (1998); Kandori et al. (1993); and Macy (1997) and the surveys in Kandori (1997) and Gintis (2000).

Institutionalism's recognition of the evolutionary and undesigned nature of institutional development and New Institutionalism's concern with intentionality. Unlike the intentionally created perspective and like the evolutionary perspective, the historical-process perspective seeks to account for an institution's emergence, stability, and change by exploring the forces that render it an equilibrium. Unlike the evolutionary perspective, it places the social level (institutional elements) at the center of the analysis and considers processes of change and the microfoundations of inertia to be endogenous. It thus extends the study of the intertemporal relationships among institutions to situations that cannot be captured in political economy or evolutionary models.

By bridging the gap between the evolutionary and intentionally created perspectives, the historical-process perspective contributes to the development of both. In contrast to the evolutionary perspective, which takes institutional inertia, mutation, and experimentations as exogenous, the perspective developed here explores the micro-foundations of institutional inertia and captures the fact that mutation and experimentation depend on existing institutions. It enriches the intentionally created perspective by recognizing that individuals look forward through the prism implied by past institutions, that an institution's equilibrium nature limits its response to functional needs, and that different institutions imply distinct institutional trajectories.

Perhaps more important, the historical-process perspective presents a new direction in social-sciences-oriented historical research. This research has long followed in the footsteps of such giants as Marx and Malthus in seeking a deterministic theory of history. The flow of history reflects the shackles of such inescapable forces as geography, class struggle, and demography. The historical-process perspective suggests an alternative: history unfolds based on the nondeterministic impact of past institutions on outcomes in general and institutional dynamics in particular.

Although the analytical development of the argument is still in its preliminary stages, Part III introduces the broad argument and empirically demonstrates its merit. Chapter 6 presents a theory of endogenous institutional change. Chapter 7 discusses the influence of past institutions on the direction of institutional change. Chapters 8 and 9 present empirical studies of institutional

dynamics, showing how different societies embarked on distinct institutional trajectories. These studies do not capture all the aspects of the argument advanced in the theoretical chapters, but they illustrate various aspects of it.

Specifically, Chapter 8 focuses on the dynamics of the institutional foundation of polities, examining the Republic of Genoa. Understanding political order and disorder requires departing from a long tradition of studying these issues while focusing on the relationships between political institutions defined as rules governing political decision making, political order, and economic prosperity (see, e.g., Przeworski 1991). The analysis here considers polities as self-enforcing institutions whose details generate the behaviors leading to political order, disorder, and economic outcomes. Rules governing political decision making are only one component of these institutions. Understanding political order, disorder, and its impact on the economy requires studying the polity as a self-enforcing institution.

Chapter 9 focuses on the dynamics of economic and social institutions. It compares the organizational, contractual, and institutional development of the Maghribi and Genoese traders. Economists often assume that such developments are influenced by efficiency considerations reflecting, in particular, attempts to reduce transaction costs (O.Williamson 1985). The comparative analysis of developments in these two societies, however, establishes the importance of past institutional elements in directing them. Furthermore, these distinct institutional elements reflect cultural influence on institutional selection. Initial cultural and social factors influence institutional selection, integrate into the resulting institutions reproduced by them, and thereby exert a lasting influence on institutional, organizational, and contractual development.

Chapter 6 A Theory of Endogenous Institutional Change

A prerequisite to studying endogenous institutional change is recognizing the mechanisms that causes institutions to persist in the absence of environmental changes and to exhibit stability despite environmental changes. Sociologists such as Berger and Luckmann (1967), Searle (1995), and Giddens (1997) have long noted the importance of studying the mechanisms causing an endogenous institution to persist once it has prevailed. But sociology has not offered a satisfactory analytical framework with which to study the phenomenon. As Scott notes, "The persistence of institutions, once created, is an understudied phenomenon [in sociology]. The conventional term for persistence—inertia—seems on reflection to be too passive and nonproblematic to be an accurate aid to guide studies on this topic" (1995, p. 90; see also DiMaggio and Powell 1991a, p. 25; Thelen 1999, p. 397).

In economics the study of institutional persistence is usually referred to as the study of institutional path dependence (North 1990; David 1994; Greif 1994a). The idea of path dependence was originally developed to study technology (David 1985; Arthur 1988, 1994). It postulates that "the present state of arrangements" requires examining the "originating context or set of circumstances and [the] sequence of connecting events that allow the hand of the past to exert a continuing influence upon the shape of the present" (David 1994, p. 206).

The game-theoretic analytical framework and the view of institutions developed in the previous chapters highlight a particular mechanism for institutional persistence. In situations in which institutions generate behavior, beliefs motivate it, and observed behavior confirms the relevance of these beliefs. Taken together, self-enforcing (and reproducing) beliefs and behavior are in a steady-state equilibrium. The observed behavior reproduces the beliefs that generated it, because it confirms each individual belief that others will behave in a particular manner, and given these beliefs, it is optimal for each individual to do so. By revealing which beliefs and behavior can be self-enforcing in a given environment, the game-theoretic perspective highlights the limit of this mechanism. It exposes which exogenous change would cause the current behavior to no longer be self-enforcing and hence to change.

Studying endogenous institutional change, however, seems particularly difficult when institutions are viewed as equilibrium phenomena. In an institution, each player's behavior is a best response. The seemingly inescapable conclusion is that change in a self-enforcing institution must have an exogenous origin, because no one has an incentive to deviate from the behavior associated with the institution. As P. Hall and Taylor note, studying institutions as equilibria "embroils such analysis in a contradiction. One implication of this approach is that the starting point from which institutions are to be created is itself likely to reflect a Nash equilibrium. Thus it is not clear why the actors would agree to change in existing institutions" (1996, p. 953). Endogenous institutional change appears, then, to be a contradiction in terms.[5] Indeed, the analysis of institutional change using game theory has concentrated mainly on the dynamics triggered by changes in parameters exogenous to the institutions under study.

In this chapter I argue that the equilibrium approach can be integrated with the study of endogenous institutional change. Recognizing the distinction between institutions and game-theoretic equilibria allows two related concepts to be introduced: *quasi-parameters* and *institutional reinforcement*. Before discussing these concepts, it is important to note the distinction between parameters and variables in a game-theoretic framework. Parameters are exogenous to the game under consideration. If they change, the implied new equilibrium set needs to be studied. In contrast, variables are determined endogenously as outcomes in the game. Institutional analysis using the game-theoretic framework typically concentrates on a single transaction (e.g., abusing or protecting property rights by a ruler) and examines as variables possible related self-enforcing behavior (e.g., security of property rights) for a given set of parameters.

In contrast, this chapter asserts that it is conceptually sound and analytically tractable to consider some aspects of a situation as parameters when studying self-enforceability but as

---

[5] Although this criticism of the game-theoretic contribution is fundamentally fair, it should be noted that ex ante creation of institutional arrangements can be predicated on variables that are not realized until later. Once such a realization occurs, the institution can change as part of a dynamic equilibrium (see Muthoo and Shepsle 2003 for an example). The discussion in this chapter of stability in the face of parametric shifts notes that it is appropriate and realistic to model institutions when the long- term implications of a shift in variables are not foreseen ex ante.

variables subject to change when studying institutional dynamics. It is appropriate to inquire whether the institution, analyzed as a game-theoretic equilibrium, endogenously affects aspects of the situation apart from behavior in the transaction under consideration. The argument advanced here is that some such aspects should be considered as parametric in studying self-enforceability in the short run but as endogenously determined—and thus variable—in the long run. Parameters that are endogenously changed in this manner and with this effect are referred to here as *quasi-parameters.* Marginal changes in quasi-parameters do not lead to a change in the behavior and expected behavior associated with this institution.

Equilibrium analysis fosters the study of quasi-parameters by making explicit the factors that make a particular behavior an equilibrium. The distinction between a parameter, a variable, and a quasi-parameter is not rigid; it is based on empirical observation. If self-enforcing outcomes affect the values of one or more parameters supporting the observed equilibrium in a manner that would lead only to long-term behavioral change, these parameters are best reclassified as quasi-parameters.

An institution is reinforcing when the behavior and processes it entails, through their impact on quasi-parameters, increase the range of parameter values (and thus situations) in which the institution is self-enforcing. If an institution reinforces itself, more individuals in more situations will find it best to adhere to the behavior associated with it.[6] When they are self-reinforcing, exogenous changes in the underlying situation that otherwise would have led an institution to change fail to have this effect. An institution would be self-enforcing for a wider range of parameters. But such reinforcing processes can fail to occur. The processes an institution entails can undermine the extent to which the associated behavior is self-enforcing. The behavior an institution entails can cultivate the seeds of its own demise. Whether this change is gradual or sudden, marginal or comprehensive, depends on the nature of these processes.

[6] More specifically, any combination of more individuals in the same situation and the same number of individuals in more situations.

Considering endogenous institutional change as reflecting undermining processes ignores the impact of institutions on the incentive to invent or adopt new institutional elements or to bring about new situations. These important issues are left to the next chapter.[7]

Historical Institutionalism in political science represents the line of research that particularly focuses on institutional change (see P.Hall and Taylor 1996; Thelen 1999; Pierson and Skocpol 2002). It stresses the importance of historical processes in shaping institutions but offers no theory able to study the interrelationships among stability, processes, and change. As Pierson (2000, p. 266) notes, an important obstacle for furthering Historical Institutionalism has been that institutional changes "are usually attributed, often ex post, to 'exogenous shocks.' We should expect, however, that these change points often occur when new conditions disrupt or overwhelm the specific mechanisms that previously reproduced the existing [behavior]." Bridging the game-theoretic and historical perspectives—by examining the relationships between factors implying that an institution is self-enforcing, the processes this institution implies, and the implications of these processes on the institution's self-enforceability—enriches both perspectives (see Greif and Laitin 2004 for a discussion of the relationship between Historical Institutionalism and the perspective developed here).

In this chapter, sections 6.1 and 6.2 examine institutional persistence and stability. Section 6.3 introduces the concepts of quasi-parameters and reinforcement. Section 6.4 illustrates how self-enforcing institutions can be either self-reinforcing or self-destroying by studying political institutions in early modern Genoa and Venice. Section 6.5 presents a model of institutional reinforcement. Section 6.6 focuses on reputation-based institutions and explains why institutions may exhibit a "life cycle" in which they are first reinforced and then undermined. Section 6.7 considers the argument's further development.

6.1 Persistence

[7] Similarly, for simplicity of presentation, this chapter focuses mainly on beliefs rather than on norms. Extending the argument to the case of norms is possible, however, building on the discussion in section 5.3.

As already noted in Chapter 5, for an established institution to persist through time, it must be *reproduced*. An institution is reproduced when the rules and beliefs that enable, guide and motivate an individual's actions are not being refuted by observed behavior or outcomes. Thus, observed behavior and outcomes confirm the rules and beliefs that enabled, guided, and motivated the original behavior, as expectations are consistent with outcomes.

D. Lewis (1969, pp. 41–2) beautifully expresses the idea of the reproduction of beliefs by behavior in equilibria: "Each new action in conformity to the regularity [of behavior associated with this equilibrium] adds to our experience of general conformity," he writes. "Our experience of general conformity in the past leads us, by force of precedent, to expect a like conformity in the future. And so it goes—we're here because we're here because we're here because we're here. Once the process gets started, we have a metastable, self-perpetuating system of preferences, expectations, and actions capable of persisting indefinitely." The structure generates behavior that, because it is self-enforcing, reproduces that structure.

This mechanism for persistence rests on intuitively appealing propositions. Individuals are forward-looking: they look before they leap and take into account what others are likely to do. They are also retrospective, evaluating their beliefs based on observable outcomes.[8] This mechanism for persistence is captured by the Nash condition, which requires each individual to hold the correct beliefs about others' behavior. (See Appendix A and Chapter 5.) Any institution that is self-enforcing in the Nash sense also reproduces itself by the behavior it generates.

The historical examples of the previous chapters illustrate the relevance of the causal mechanism for institutional persistence that the Nash restriction captures. The persistence of the Maghribis' coalition, for example, reflects the self-enforceability of correct behavioral beliefs and behavior. Each trader's best response to the belief that everyone will follow a particular behavioral rule was to follow it as well. The observed behavior of hiring only member agents and honesty, in turn, reproduced (confirmed) these beliefs.

Game theory thus captures the conditions under which, and the mechanism by which, the structure—commonly known rules and beliefs—generates behavior that reproduces this structure.

---

[8] Because this mechanism regards the relationships between beliefs and behavior, it is applicable to all transactions (economic, information-sharing, coercive, legal, political, social).

The distinction between game-theoretic and institutional analysis, however, is worth recognizing. First, game-theoretic analysis assumes that players have common knowledge of the rules of the game; institutional analysis recognizes that individuals play against the (institutionalized) rules and learn about various aspects of the situation through social rules, others' behavior, and similar observable outcomes. As I show later, this implies that some individuals may not recognize underlying changes in various aspects of the situation and will therefore not change their behavior accordingly. In such cases, institutions can and often do persist despite parametric changes.

This mechanism for institutional persistence also contributes to the persistence of what is often referred to as a society's cultural and social (organizational) features. Institutionalized rules, beliefs, norms, and organizations are components of institutions that generate behavior. At the same time, they are also part of the society's cultural and social features, because they imply social positions, are embodied in individuals' preferences, and constitute internalized and other beliefs that are commonly known societal features. The overlap between institutional features on the one hand and cultural and social features on the other implies that the described mechanism for institutional persistence contributes to the persistence of a society's cultural and social features. Cultural and social features can be maintained, however, based on other mechanisms such as the transmission of norms through socialization and the desire of individuals to maintain their social identity.

The Maghribis traders' social structure—the Maghribi traders' group—was an integral part of an institution that fostered the welfare of the groups' members. The different behavior toward members and nonmembers that this institution implied reproduced this distinct social identity. The merchant guild organizations were reproduced in a similar manner. This reproduction process implies that the endogenous processes that render a particular institution no longer self-enforcing also imply that its overlapping cultural and organizational features can no longer be reproduced by the behavior the related institution entails.

6.2 Stability in the Face of an Endogenous Parametric Shift

Game-theoretic analyses of institutions have traditionally focused on studying the relationships between the rules of the game and equilibrium behavior—cooperation, wars, political mobilization, social unrest—in the transactions captured in the game. Such analyses make explicit the dependency of possible equilibria, and hence institutions, on various parameters (such as payoffs from various actions, time discount factors, risk preferences, wealth, and the number of players) of the underlying game. The framework highlights the conditions under which an exogenous change in parameters will render an institution no longer self-enforcing.

Focusing on regularities of behavior in a particular transaction for a given set of parameters diverts attention from other possible ramifications of an institution that go beyond this behavior. Institutions influence factors—such as wealth, identity, ability, knowledge, beliefs, residential distribution, and occupational specialization—that are usually assumed as parametric in the rules of the game. Although it may not be possible to prove that institutions generally have such ramifications, it is difficult to think of any institution that in the long run does not have implications beyond the behavior in the transaction it governs. In the game-theoretic framework, such influence implies a dynamic adjustment of variables that, had this influence been ignored, would have been considered parameters in the stage game (i.e., a game repeated every period; see Appendix A).

In the game-theoretic framework, such changes would not necessarily lead to behavioral change. The Folk theorem of repeated games (presented in appendix A) exemplifies the general game-theoretic insight that, for a given parameter set, a multiplicity of equilibria usually exists. The theorem also highlights a corollary to this insight: a particular equilibrium can usually be sustained over a broad range of parameters. If a strategy combination is an equilibrium, it is usually an equilibrium in some parameter *set*. Game theorists have long recognized that game theory does not predict behavioral change following a parametric change. Moriguchi (1998) refers to the set of parameters in which a particular strategy set is an equilibrium—and hence the associated institution can prevail—as "institutional support."

Indeed, there are good reasons for individuals to continue to follow past patterns of behavior even under conditions of marginal parametric change. This is the case for various

interrelated reasons, such as knowledge and coordination, which were touched on in Chapter 5. Other reasons, such as attention and habit, are introduced here.

6.2.1 Knowledge and Playing against the Rules

In Chapter 5, I argued that institutionalized rules provide the cognitive, coordinative, and informational basis for behavior at the individual level. Institutionalized rules of behavior aggregate cognition, knowledge, and information in a compressed form and direct individuals to play an equilibrium strategy in the game thereby constructed. Individuals play against the (institutionalized) rules rather than against the commonly known rules of the game.

Hence past behavior can reign, and an individual will continue to follow past institutionalized rules of behavior despite marginal parametric changes. This outcome occurs because institutionalized rules learned in the past convey these cognitive models, provide aggregate information, and guide behavior. As long as the behaviors of others (the causal underpinnings of which one may not understand) do not reflect that these models are mistaken or that the parameters have changed, an individual will not change his behavior if it is still in his best interest to follow it while responding to the cognitive and informational content of the prevailing institutionalized rules. In other words, the fact that actors play against the rules implies that changes in various aspects that are incorporated into the rules of the game influence behavior only when those who observe them reveal them through their behavior.[9] If they do not, behavior continues to reproduce beliefs, and the institution persists.

6.2.2 Coordination

Schelling's (1960) seminal work on focal points highlights the importance of coordination in choosing behavior in strategic situations characterized by multiple equilibria. The related argument made here is that the need for coordination implies that individuals continue to follow past patterns of behavior, even under conditions of observed marginal parametric change. They do so because they face a situation in which rationality alone is insufficient to select a behavior

[9] Chapter 3 discusses models of incomplete information that explore one's motivation and ability to reveal his information to others. For a general discussion, see Fudenberg and Tirole (1991).

(because of multiple self-enforcing outcomes). They therefore rely on institutionalized rules to guide them. Under these circumstances, behavioral rules learned in the past are the best predictor of future behavior, even when some individuals and organizations have the ability to coordinate on new behavior. For many reasons, such coordination may fail to transpire even when it is beneficial. Sunk costs associated with coordinating change, free-rider problems, distributional issues, uncertainties, limited understanding of alternatives, and asymmetric information may hinder coordination on new behavior. In the terminology developed here, the need to coordinate on one out of many possible behaviors implies that even observed marginal changes in the rules of the game are not likely to cause behavioral changes, because past behavior constitutes a focal point.

6.2.3 Attention

What an individual sees, knows, and understands in a given situation reflects the amount of attention he devotes to the task. Attention is a scarce resource (Simon 1976); institutionalized rules allow individuals to choose behavior in complicated situations while devoting their limited attention to decision making in noninstitutionalized situations. People do not consider their optimal response to every choice they make in life (DiMaggio and Powell 1991a). In particular, they do not consider such responses in situations in which institutions guide their behavior. In such situations, parametric shifts that might have been noticed if more attention had been devoted to observing them may go unnoticed, contributing to the lack of behavioral change. Moreover, those who observe the parametric shift and can bring it to the attention of others may not have the incentive to do so. Limited attention capacity implies that even potentially observable changes in the rules of the game may go unnoticed and hence not influence behavior.

6.2.4 Habit and Scarce Cognitive Resources

Judgment and habit are interrelated in influencing behavior (Margolis 1987, p. 29).[10] But once a particular pattern of behavior has been institutionalized, individuals tend to rely more on habits and routines than on reason and calculations. We follow institutionalized behavior habitually because of the scarcity of cognitive resources (see Clark 1997a, 1997b; R.Nelson and Winter 1982; R.Nelson 1995; March and Olsen 1989). Habit enables people to devote scare cognitive resources to other tasks. When individuals are guided by habit and routine and rely less on judgment, past behavior reigns despite marginal parametric changes.

6.3 Quasi-Parameters and Reinforcement

Many features that are usually taken as parameters in the repeated-game formulation share two properties: they can gradually be altered by the implications of the institution under study, and marginal changes to them will not necessarily cause the behavior associated with the institution to change. These features do not cause the behavior associated with the institution to change because, ex ante, people do not recognize, anticipate, directly observe, understand, or pay attention to the changes in these features and the ramifications of those changes for the institution. Even when this is not the case, because of ex post coordination problems, these changes do not cause the behavior associated with the institution to change. These features are neither parameters (as they are endogenously changed) nor variables (as they do not directly condition behavior); they are quasi-parameters. Because the actors do not recognize changes in quasi-parameters or their implications, quasi-parameters must be considered as parametric —exogenous and fixed—in studying the self-enforcing property of an institution in the short run but as endogenous and variable when studying the same institutions in the long run.[11]

---

[10] The analysis of habit and institutions can be traced back at least to Simon (1976). Berger (1977) and Kuran (1993) argue that institutionalized behavior has become the social equivalent of an instinct. Margolis (1994) and G. Hodgson (1998) identify habits with institutions.

[11] Institutional elements and their attributes (e.g., the size of a community) can be quasi-parameters. An institutional element is part of a system that generates behavior, implying that each of the notions—an institutional element and a quasi-parameter—highlights a distinct characteristic of a social factor.

Changes in quasi-parameters implied by an institution can reinforce or undermine that institution. An institution reinforces itself when, over time, the changes in quasi-parameters it entails imply that the associated behavior is self-enforcing in a larger set of situations—a larger set of other parameters—than would otherwise have been the case. A self-enforcing institution that reinforces itself is a *self-reinforcing* institution. A self-enforcing institution can also undermine itself when the changes in the quasi-parameters it entails imply that the associated behavior will be self-enforcing in a smaller set of situations.

Central to endogenous institutional changes are therefore the dynamics of self-enforcing beliefs and the associated behavior. A change in beliefs constitutes an institutional change; it occurs when the associated behavior is no longer self-enforcing, leading individuals to act in a manner that does not reproduce the associated beliefs.[12] Undermining processes can lead previously self-enforcing behavior to cease being so, leading to institutional change. A sufficient condition for endogenous institutional change is that the institution's implications constantly undermine the associated behavior. Conversely, a necessary condition for an institution to prevail over time is that the range of situations in which the associated behavior is self-enforcing does not decrease over time: the institution's behavioral implications have to reinforce it, at least weakly. Hence unless an institution is (weakly) self-reinforced, eventually the behavior associated with it will not be self-enforcing, and endogenous institutional change will occur.

Considering reinforcement highlights the importance of another, indirect way in which an institution endogenously influences its change: by affecting the magnitude and nature of the exogenous shocks necessary to cause the beliefs and behavior associated with the institution to change. When an institution reinforces itself, the behavior associated with it does not change, but the reinforced institution is nevertheless more robust than before. The behavior associated with it becomes self-enforcing in situations in which it previously would not have been. Reinforcement

[12] The focus here is on endogenous institutional change due to self-reinforcement and undermining, but the observations about the nature of institutions, institutionalized rules, and beliefs allow us to extend the analysis easily to address related issues, such as intentional coordinated action to change others' beliefs, to draw attention to change, to coordinate actions by some to influence others' optimal behavior, and to establish organizations that foster or halt reinforcement or undermining. Some of these issues are discussed later.

implies institutional hysteresis; the institution will be self-enforcing in a situations in which, prior to its reinforcement, it would not have been. The opposite holds in the case of an institution that undermines itself. By reinforcing or undermining itself, an institution indirectly influences its change by determining the magnitude of an external change in parameters required to render behavior associated with it no longer self-enforcing.

Institutions can change due to endogenous processes, exogenous shocks, or combinations of both. The mechanism that brings about institutional change once the behavior associated with an institution is no longer self-enforcing depends on the nature of the quasi-parameters that delimit self-reinforcement. If these changes in quasi-parameters are unobservable, uncertain, and unrecognizable, the mechanism of institutional change is likely to reflect individuals' willingness to experiment and risk deviating from past behavior or the emergence of individuals with better knowledge of the situation, who, through their behavior, reveal a new institutional equilibrium.[13] In either case, learning is slow, and it may take a long time for self-undermining to be reflected in new behavior.

When the undermining that leads to the institutional change is not foreseen ex ante but many individuals recognize ex post that following past behavior is no longer optimal, the change will be manifest by the sudden abandonment of past behavior.[14] Institutional change can thus be characterized by punctuated equilibria (Gould and Eldrege 1977; Krasner 1984; Aoki 2001), in which change is actually evolutionary but appears to be abrupt. Such abrupt change is typically associated with a crisis that reveals that the previous behavior is no longer an equilibrium.

An institution can also cease to be self-enforcing due to changes in quasi-parameters that are observable and whose importance is understood. When the impending change in behavior becomes progressively more recognizable, decision makers will realize that past behavior is

---

[13] Game theory highlights the importance of uncertainty in these processes. If the eventual collapse of the institution is known and expected to prevail at a particular time, the transaction has to be modeled as a finite game. The set of behaviors that is self-enforcing in these games is much smaller than the behaviors that can prevail in an infinitely repeated game. If the eventual collapse is not expected or its timing is uncertain, the equilibrium set is much larger (see Appendix A).

[14] Gradual processes of institutional change are discussed later in this chapter and in Chapter 7.

becoming less self-enforcing, and the mechanism directly leading to institutional change will be intentional and is likely to be gradual. Alternative behaviors, specification of new rules through collective decision making, and intentional introduction of organizations are common manifestations of this mechanism. Such institutional change often manifests itself in intentional reinforcement—the preemptive introduction of reinforcing institutional elements—which is likely to occur gradually. Institutional change in this case will take the form of restoring the prechange behavior but supporting it with different institutional elements. We have seen just that in considering the organizational evolution of the merchant guild institution.

Like intentional reinforcement in the face of anticipated self-undermining, the prevalence of a particular institution can induce coordinated actions aimed at undermining it and instituting other self-enforcing behavior. Such coordinated undermining reflects the fact that, although no individual dissatisfied with the prevailing institution can change it, individuals acting collectively may be able to do so.[15] They can undermine it by, for example, aggregating their resources and using them to increase the payoffs others receive from following the behavior they want to institute. Resources are needed, because the institutionalized behavior is everyone's best response, and inducing someone to adopt a so-far noninstitutionalized behavior requires changing motivation (by, e.g., changing beliefs regarding its consequences). Once the behavior of a sufficiently large number of people has shifted to a new self-enforcing behavior, the best response for all others is to adopt the behavior as well. The previous institution has been undermined, and a new behavior becomes institutionalized. Once it is, it may no longer be necessary to devote resources to inducing this behavior.

6.4 Self-Reinforcement: A Tale of Two Cities

To illustrate this dynamic approach to institutional change, I examine the experiences of late medieval Venice and Genoa, analyzing the political regime of each city as an institution made up

[15] I discuss the role of leadership in institutional change in Chapter 7.

of the following elements: the organization of the governing structures; the rules for choosing leadership positions and behavior; and the norms, rules, and beliefs shared by their citizens.[16]

The residents of the settlements around the Venetian lagoon established Venice as a political unit in 697; residents of Genoa organized themselves into a commune around 1096. By the mid-fourteenth century, Venice and Genoa had become the two most commercially successful maritime city-states in the Italian peninsula.

The rise of both cities reflects opportunities for commercial expansion made possible by the naval and military decline of Muslim and Byzantine forces around the Mediterranean, particularly during the eleventh century. During this century, however, both cities found themselves in a political vacuum, as neither the Byzantine Empire (which claimed sovereignty over Venice) nor the Holy Roman Empire (which claimed sovereignty over Genoa) was in a position to interfere in local political developments.

As result of the decline in central authority, clans and lineages became the prominent unit of social organization in both cities (D.Hughes 1978). As Herlihy notes, "The corporate or consortial family was better able than the nuclear household to defend its wealth and status [increasing] family solidarity, at least among the aristocratic classes" (1969, p. 178). In both Genoa and Venice, the strongest clans agreed to cooperate politically in order to advance their economic interests.[17] The resulting political institutions governed a particular transaction: motivating members of the cities' strong clans and families to delegate decision-making power and resources in return for political order and the economic benefits of collective action.

---

[16] For a general discussion of Venetian and Genoese history, see Lane (1973) and Epstein (1996) respectively. The analysis here builds mainly on Greif (1995, 1998c). For an illuminating analysis of the Venetian polity as a self-enforcing institution, see González de Lara (2004).

[17] An agreement for interclan cooperation does not imply that clans were unwilling to use force against one another to advance their own interests. Indeed, the historical records are rich with evidence indicating that moral considerations—internalized constraints—were not sufficient to deter one Genoese clan from using force against another and that clans aspired to achieve political dominance (Greif 1998c; Tabacco 1989). Genoa's two dominant viscountal clans were a product of the feudal world, in which the objective was to become a lord within one's domain. At the same time, the tight internal organization and military and economic resources of these clans were such that, for each, gaining control over a city was not out of reach.

The political organizations of Genoa and Venice were seemingly identical. Both cities were governed by oligarchies, their political leaders were, de jure, elected by the citizenry as a whole and subject to the law. At the top of Venice's political system was a doge and the Ducal Council. Genoa was governed initially by consuls and, after 1194, by one or more executives—called the *podestà* (power)—and a council of rectors. The political institutions that prevailed in Venice and Genoa from the late eleventh century were both able to support interclan cooperation that initially fostered commercial expansion and political order.

Despite these similarities, the histories of the two cities differ greatly. Venice was able to maintain political order in a changing economic environment and to mobilize resources to sustain its economic prosperity even following the decline of its trade with the Far East. Throughout its history, its members' social attachments to the clan structure gradually declined. In contrast, in Genoa political order often broke down, contributing to the city's economic decline, and the social and political importance of clans grew.

How can we account for these different trajectories in cities that faced similar initial conditions, outside opportunities, and basic political structures? To understand these histories and their long-term implications, I examine these cities' institutions. The origins of Genoa's and Venice's two distinct institutions are not the focus of the analysis. Yet the institutional differences that account for Venice's relative success probably reflect the institutional heritage of the post of the doge, its less unequal initial distribution of interclan military might and wealth, and a series of able leaders who coordinated and developed elements of Venice's institutions.

Both Genoa and Venice initially developed political regimes that were sufficiently self-enforcing to sustain interclan cooperation and economic prosperity. But Genoa's institutions were self-undermining, whereas Venice's were self-reinforcing. To develop this argument, I consider quasi-parameters, such as the wealth of the cities, the strength of the *popoli* (roughly speaking, the nonnobles), and the social identities of the clans. Understanding these cities' subsequent histories requires examining the dynamics of these quasi-parameters in two different institutional equilibria. Changes in the quasi-parameters in Genoa undermined political order, making its institutions sensitive to relatively small exogenous shifts in clans' strength, trading

opportunities, and level of external threat. Changes in the quasi-parameters had the opposite effect in Venice.

6.4.1 Genoa

During its first hundred years (1096–1194), elected consuls were Genoa's political, administrative, and military leaders. These consuls were representatives of the main Genoese clans (D. Hughes 1978, pp. 112–13). Control of the consulate enabled clans to gain economically from the city's resources and power. The behavior of these consuls and the clans they represented was guided by the belief that clans would challenge one another militarily if the opportunity arose to gain political dominance over the city. The self-enforcing institution that governed the clans' interrelationships was thereby based on mutual deterrence: each of Genoa's two main clans expected the other to use its military might to gain political and economic dominance over the city, but each was deterred from doing so by the other's military strength. Hence each of Genoa's two main clans was motivated to mobilize its resources for interclan cooperation to advance Genoa's economy, albeit only to the extent to which its ability to deter other clans from militarily challenging it was not weakened.

Initially, the relatively high gains from the joint mobilization of resources implied that interclan rivalry did not hinder interclan cooperation. But because interclan cooperation advanced Genoa's economic prosperity (an endogenous change in a quasi-parameter), it rendered political control over the city a more rewarding objective and intensified the competition over political and economic dominance in the city. Fearing that any temporary decline in their relative power would constitute an opportunity that the other clan would exploit, clans became engaged in an "arms race," which led to yet other endogenous changes in quasi-parameters: the purchase of land, which clans fortified to dominate particular quarters; the establishment of patronage networks; and the socialization of clan members to internalize loyalty to the clan and the norm of revenge to protect clan honor.

A foreign threat constituted an exogenous shift in parameters that sustained interclan cooperation. For a period after 1154, attempts by the emperor Frederick Barbarossa to regain de facto control over northern Italy rendered mutual deterrence self-enforcing in a larger set of

parameters. This external threat did not alter clan members' beliefs about what other clans would do if the threat receded, but because the clans expected the threat to last, every clan had a reduced incentive to challenge another clan militarily. The result was that the Genoese clans jointly mobilized their resources, acquired overseas commercial possessions, and expanded commercially, as Genoa's economic structure was transformed into one based on long-distance commerce.

This commercial expansion and structural transformation undermined interclan mutual deterrence by making it self-enforcing for a smaller range of parameters. Greater economic prosperity, which increased the gains from controlling the city, implied a smaller set of parameters for which mutual deterrence was self-enforcing in the absence of an external threat.

In 1164 civil wars in Germany diverted the emperor's attention from Italy. As a result, the level of external threat facing Genoa substantially declined, possibly returning to its pre-1154 level. But the quasi-parameter of wealth was now higher than it was before, and with beliefs remaining stable, the previous mutual-deterrence equilibrium among the clans was no longer self-enforcing. The commune sank into a lengthy civil war, during which various clans gained the upper hand for a time, only to be challenged when exogenous conditions changed. As a twelfth-century Genoese chronicler observed, "Civil discords and hateful conspiracies and divisions had risen in the city on account of the mutual envy of the many men who greatly wished to hold office as consuls of the commune" (*Annali* 1190, vol. II, pp. 219–20). The fighting was particularly devastating between 1189 and 1194, when it endangered the city's very existence.

These events reflect more than just the shift in exogenous conditions. They reflect the fact that endogenous changes—increasing commercialization and prosperity, the clans' past investments in military ability and patronage, and perhaps the fomenting of individuals' identities as clan members—made Genoa's institution self-enforcing for a smaller set of parameters. The city that was peaceful despite the absence of a threat by an emperor before 1154 became embroiled in a civil war during the emperor's absence after 1164. An exogenous situation that previously would not have led to the collapse of Genoa's institution now had a devastating effect.

In 1194 the Holy Roman Emperor, who needed the assistance of Genoa's navy, had an interest in ending the civil war. By promises of rewards and threats of war, the emperor induced the Genoese clans to agree to alter Genoa's political institutions by introducing a self-enforcing organization that restored interclan mutual deterrence and cooperation.

At the center of Genoa's new institution was a non-Genoese *podestà*. The *podestà* was selected by a committee of representatives from the city's neighborhoods, a committee that was large enough that no clan dominated it. Hired for a year to serve as Genoa's military leader, judge, and administrator, the *podestà* was supported by the soldiers and judges he brought with him.

The *podestà* and his military contingent fostered the clans' ability to cooperate by creating a military balance among them. The threat of intervention by the *podestà* deterred each clan from attacking the other to gain control over the city. Because the *podestà* was paid at the end of his term, the threat was credible, because if a clan took control of the city, there was no reason why it should pay the *podestà.* This reward scheme also made it in the *podestà's* interest not to alter fundamentally the balance of power among the factions. The *podestà* could thus credibly commit to be impartial and to retaliate only against individuals who broke the law rather than against an entire clan.

For a while, the *podesteria* fostered interclan cooperation—and thus political stability and economic growth. It was a self-enforcing institution, as the self-enforcing belief in the futility of gaining political dominance by using force deterred clans from trying. The belief that all clans could gain from cooperation without risking their economic position through military confrontation also motivated cooperation.

Yet, like the consular system that prevailed prior to the *podesteria*, the *podesteria* was not reinforcing—indeed, it contained the seeds of its own destruction. Specifically, because the *podesteria* was based on a balance of military strength among the clans and each clan wanted to be militarily prepared in case of need, it contained but did not eliminate interclan rivalry. Each clan was still motivated to strengthen itself militarily with respect to the others, and clan members' main identification was still with their clan and not the city.

The creation of the *alberghi* and the rise of the *popolo* as a faction during this period were further manifestations of the lack of reinforcement of the institutional equilibrium. *Alberghi* were clanlike social structures whose purpose was to strengthen consorterial ties among members of various families through a formal contract and the assumption of a common surname, usually that of the *albergo's* most powerful clan (D.Hughes 1978, pp. 129–30). By the fifteenth century, about thirty *alberghi*, each containing five to fifteen lineages, dominated political and economic life in Genoa. At the same time, each clan's attempt to develop a patronage network and to provide access for all city residents to Genoa's lucrative overseas trade implied that over time the *popolo* acquired the resources, organization, and recognition of their common interests to form a political faction that disrupted the noble-controlled equilibrium.

Under the *podesteria,* peace was maintained. But Genoa's institutions motivated clans to establish patronage networks (thereby mobilizing the *popoli*), to indoctrinate their members to internalize the norms of revenge and adopt identities (through the *alberghi*) as clan members, to fortify their residences, and to acquire the military ability to attack other clans. These changes did not render the *podesteria* ineffective in the short run; it remained self-enforcing. But over time these changes caused Genoa's political structure to become self-enforcing in a smaller range of situations, leading to its eventual demise. In the long run, a *podestà* could not constrain the balance of power incentives among Genoa's rival clans, and the system collapsed.

6.4.2 Venice

The early history of Venice parallels that of Genoa. After an initial period of interclan cooperation, interclan rivalry developed, with the goal of capturing the office of the doge (Lane 1973; Norwich 1989). Originally, the doge was a Byzantine official, but shortly after Venice was established (in 679), the post became that of an elected monarch with judicial, executive, and legislative powers . For the next few hundred years, clans fought for control over the doge's post. As in Genoa, economic cooperation was hindered by the lack of an institution able to contain interclan rivalry.

Changes around the Mediterranean increased the cost of such confrontations. Toward the end of the eleventh century, the decline of Byzantine naval power increased the gains to the

Venetians of forming a political institution that enabled cooperation. They responded by establishing a new self-enforcing institution that this opportunity made possible (self-enforcing). At its center was the belief that every clan would fight against a renegade clan that attempted to gain political dominance over the city and its economic resources.[18] This belief and the behavior it fostered may have helped forge a common Venetian identity that reinforced this belief. In any case, a set of institutionalized rules guided the behavior of the Venetians toward this self-enforcing belief and generated the conditions required for these beliefs to be self-enforcing. The rules limited the doge's power to distribute economic and political rents, curtailed the clans' ability to influence the outcome of the election of a doge (or any other officer), established tight administrative control over gains from interclan political cooperation, and allocated these rents among all the important Venetian clans so that all had a share in them regardless of clan affiliation. This allocative rule did not provide clans with incentives to increase their military strength to prepare for interclan military conflict. Because these rules were being developed at a time when the Byzantine and Islamic naval powers were on the decline and cooperation was most beneficial, Venetians were able to make the most of this opportunity.[19]

Beginning in 1032, the doge's authority was limited through the establishment of advisory councils until it was de facto altered from an elected monarchy to a republican magistracy. In 1172 the Venetians, through their representative organizations, established that a doge should never act contrary to the advice of his councillors. To inhibit the ability to use a clan's political machine and popular support to influence the election, the selection of the new doge was entrusted to an official nominating committee, which was selected and formed through an elaborate process based on both lotteries and deliberation. The (partially random) process began in the Great Council, in which all adult members of the powerful clans were eligible to participate. From this council a committee of thirty was chosen by lottery and its role was to

---

[18] How these beliefs were formed remains unclear.

[19] Muslim naval power particularly declined during the eleventh century following the political disintegration of Muslim Spain, the crumbling of centralized control over North Africa, and the military conflict between the Fatimid Caliphate centered in Egypt and the `Abbasid Caliphate, centered in Baghdad.

propose a list of candidates for the post of the doge. The selection of candidates proceeded through an additional nine steps of deliberation and selection by lot until the proposed candidate was brought before the Venetian assembly for approval. The importance of clans and their patronage network was reduced by these processes and rules requiring that only one family member could be on any committee and that a delegate had to recuse himself when a relative was being considered. The process was designed to reach a quick decision thereby reducing the ability to manipulate the system.

Similar, albeit less elaborate, selection processes were used to select other officials. Their numbers were relatively large and their time in office relatively short, so that members of many clans could hold a particular office in a given period of time. Nominating committees for many posts were selected by ballot in the Great Council in a way that gave every person present an equal chance of being on a nominating committee. To prevent officials from reaping unlawful gains, the conduct of all officials (including the doge) was subject to scrutiny by committees.

The belief that each clan would join the others to confront any clan that attempted to use military power to change the rules was self-enforcing because all clans benefitted from these rules. The rules and associated beliefs were also reinforcing, because they provided clans with few incentives to invest in fortifying their residences or establishing patronage networks. By weakening the importance of clan and linking one's prospects to the city's rules and success, the system fostered norms of loyalty to the city. By weakening the clans, over time Venice's republican magistracy increased the range of situations in which this political institution was self-enforcing. This institution also prevented the endogenous formation of a political faction among the *popoli*, because the magistracy as an institution did not motivate clans to establish patronage networks that would have channeled rents from political control over Venice's overseas possessions to nonnoble clans.[20]

6.5 Formal Representation of Institutional Reinforcement

[20] This group was extended several times to absorb emerging nonnoble families. The system therefore had the flexibility required for its perpetuation.

Repeated games are games in which the same stage game is repeated each period (Appendix A). Such games would appear to be less promising for the study of institutional dynamics than dynamic games, in which the game can be changed each period. In fact, as argued earlier and evident from the success of repeated-game theory to facilitate empirical studies, the theory of repeated games captures important ways in which people view their environment and make decisions. This theory does not impose the unrealistic informational requirements or involve the computational complexities of dynamic games, which render such games unrealistically demanding as a basis for a general theory of institutional change. For these reasons, I model endogenous institutional dynamics using the framework provided by the theory of repeated games.

This section provides a formal representation of a game in which there is the possibility of an endogenous shift in one of the parameters of the game (the payoffs).[21] It illustrates how quasi-parameters and reinforcement processes can be incorporated into standard repeated-game-theoretic models. To illustrate the generality of the illustrative discussion, I relate it to the empirical analyses already discussed.

The game-theoretic framework makes explicit the parameters delineating the extent of self-enforceability of various beliefs in a game that is conditional on the relevant intertransactional linkages. Building on this framework allows us to study institutional dynamics by combining what the analyst understands about the situation—particularly regarding processes that reinforce or undermine (quasi-) parameters—with a conjecture about what decision makers understand, know, and observe.

To grasp the implications of this formulation, consider the infinitely repeated prisoners' dilemma game presented in Annex 6. In order to focus on the relationships between self-enforcing institutions and reinforcement, this model considers only one institutional element, that of shared beliefs of mutual cooperation (the outcome of the strategy (c, c) in equilibrium over

[21] The force of the argument about the importance of self-enforcing and undermining processes is not limited to the particular game structure or equilibrium refinement. It rests on the difficulties individuals normally face when having to think their way through strategic situations.

repeated play).[22] This game has four parameters: the cooperative payoff for each player (b_t), the sucker's payoff (k), the additional payoff for defecting while the other player cooperates (e), and the discount factor (δ). In this representation, b_t is a quasi-parameter.

This game differs from the standard repeated-play prisoners' dilemma model in that it allows for neutral, positive, and negative feedback from past behavior to the quasi-parameter that lead to neutral, positive, and negative self-reinforcement (undermining), respectively. In a positive feedback situation, the payoff b after any (c, c) outcome increases by ϵ for the next round of play, reinforcing the institution. In a negative feedback situation, the payoff b after any (c, c) outcome decreases by ϵ for the next round of play, undermining the institution. The cooperative payoff changes depending on the outcome in the previous round of the game. In the case of positive reinforcement, over time the range of δ for which (c, c) will be self-enforcing increases: the institution of cooperation is thus not only self-enforcing but self-reinforcing. It is an equilibrium in the short run that, in the long run, is an equilibrium for a wider range of discount factors and other parameters.

Conversely, in the case of undermining, cooperation is self-enforcing but not self-reinforcing, as the range of δ for which (c, c) is self-enforcing decreases over time. At some t in the future, cooperation will no longer be self-enforcing, and (d, d) will become the behavior associated with the new institution.

In this game, reinforcement and undermining processes do not depend on players' knowledge of the feedback mechanism. But whoever possesses this knowledge determines the institutional ramifications of these processes. Consider first a situation in which the actors are fully aware of the reinforcing (undermining process) (case 1). In this case, positive reinforcement

[22] In asserting that the players are engaged in the prisoners' dilemma game, I am asserting that particular institutional elements are or are not relevant. A legal system is implicitly assumed to exist and to be able to commit to taking particular actions in response to a prisoner's action. This implicit assumption is reflected in the game's payoff, which captures the prisoners' beliefs that cooperation reduces punishment. Potentially relevant organizations such as the Mafia are assumed not to exist. The game thus assumes away the possibility of beliefs that a prisoner who defected would be penalized by such an organization. The analysis also assumes away the possible influence of norms, such as that of honor among thieves, which the prisoners may have internalized before their arrest. Internalization of such norms would affect the prisoners' payoff from cooperating or defecting.

extends the set of parameters (δ, e, -k, b_0) in which cooperation is self-enforcing (claim 1). Cooperation would be more fragile to exogenous shocks earlier in the process. Indeed, Venice's political institution faced its most challenging moment in its early days. Alternatively, negative reinforcement reduces the set of parameters in which cooperation is possible, and cooperation, due to unraveling, would never be an equilibrium outcome.

In reality, other responses to foreseen undermining processes are often possible. The study of Genoa reflects two of them. First, cooperation led to undermining, by increasing wealth and hence the temptation to capture it. Each Genoese clan was motivated to cooperate with other clans only to the extent to which its gains from additional wealth outweighed the expected benefits of military conquest. The response to undermining was thus behavioral: ceasing cooperation while retaining the institutions of mutual deterrence.

A second possible response is organizational, altering the organizational component of the institution to restore its self-enforceability. In 1194 the mutual-deterrence equilibrium was no longer self-enforcing, but its costs to both clans increased as a result of the emperor's threat to intervene. The response was organizational: the introduction of the *podestà*, an organization designed to restore mutual deterrence and cooperation that reflected a process of learning.

In case 2 the relevant players do not recognize the reinforcing and undermining processes. In the prisoners' dilemma game, ignorance of undermining would imply cooperation for several periods until the players recognized that the situation had changed and responded by defecting. But the dynamics can take other forms, reflecting more complex situations. Even if an undermining process is recognized, the incentives implied by the self-enforcing institution may imply that players will not effectively respond to it.

Often those who observe a process of undermining have little incentive to reveal it to others. Such one-sided knowledge regarding undermining leads to the collapse of the previous institution only once the person who possesses the knowledge begins acting in a manner that reveals his knowledge. This collapse can then be followed by institutional refinement and redesign aimed at restoring a desired outcome, given new knowledge about the situation.

6.6 The Institutional Life Cycle

Institutions may have non-deterministic "life cycles" as suggested by Genoa's history. Initially, institutions tend to reinforce themselves, but undermining processes assert themselves as time passes. Initial reinforcement reflects, among other factors, the role of institutions in providing the cognitive, coordinative, and informational foundations of behavior. In the processes of institutionalization, each individual faces some uncertainty as to whether the behavior in the process of institutionalization will or will not be followed and to what effect. Basing one's actions on beliefs about what others will do is not foolproof. Others' actions are not known with certainty ex ante, and, as stressed in Chapter 5, many factors influencing other peoples' behaviors and outcomes are not directly observable. The ex ante expected value of goal-oriented behavior may be high, but these strategies could still fail ex post. When these behaviors work ex post, uncertainty is resolved—the mechanism for institutional persistence sets in—and the value of continuing to use them is higher than it was ex ante. The fact that a particular behavior led to particular results reinforces the belief that the strategy adopted by the relevant decision makers will produce the same results in the future, making it more likely to be followed.[23]

Furthermore, institutions shape individuals in ways that tend to reinforce these institutions by making the cost of deviation from the behavior these institutions generate emotionally or socially costly. Institutionalized behavior and the associated outcomes lead to reinforcing norms, senses of entitlements, identities, self-images, thinking patterns, and ideologies.[24] Regularities of behavior tend to become the normatively appropriate and fair way to behave, they gain legitimacy, lead to the development of congruent personalities, and are incorporated in individuals' identities. Once this happens, subsequent socialization further reinforces the institution. This social and psychological reinforcement implied by an institution

---

[23] One way to integrate this argument formally in the models presented here is to extend them to reflect incomplete information (see, e.g., the discussion in Fudenberg and Tirole 1993). Individuals have some beliefs over the "type" of others and hence their responses in various situations. There is thus a distinction between ex ante and ex post beliefs about their actions.

[24] K. Davis (1949); Homans (1950); Berger and Luckmann (1967); Scott (1987); March and Olsen (1989); Mead (1967 [1934]); Sugden (1989); Rabin (1993, 1994); Fudenberg and Levine (1993); G. Hodgson (1998); Kuran (1998, chapters 10-14, 1998); Ben-Ner and Putterman (1998); and Akerlof and Kranton (2000) present economic analyses of such features and processes.

tends to lead to political activities aimed at reinforcing it through laws and regulations. Similarly, those who economically benefit from existing institutions tend to have the means and influence required to pursue such activities (Olson 1982; North 1990; Mahoney 2000; Pierson 2000).[25] Finally, institutions motivate the establishment of reinforcing organizations and the acquisition of complementary capabilities, knowledge, and human and physical capital that reinforce them (Rosenberg 1982; Nelson and Winter 1982; North 1981; David 1994).

Once this initial stage of reinforcement occurs, undermining may set in, although the conditions under which this might occur is not yet clear. No general theory identifies attributes of institutions that lead to undermining.

Reputation-based institutions, however, undermine themselves when the implied behavior decreases the expected value of future rewards or sanctions (see Appendix C.) This decrease renders the beliefs that motivate the behavior associated with the institution self-enforcing in a smaller set of parameters. This is the case because, in reputation-based institutions, the fear of losing rewards or being sanctioned motivates the institutionalized behavior. If this behavior undermines these rewards and sanctions and the incentives they imply, the institutionalized behavior eventually ceases to be an equilibrium.

This mechanism through which private-order, reputation-based institutions undermine themselves is reflected in three empirical studies in this volume: the evolution of the merchant guild, discussed in Chapter 4; the undermining of Genoa's political institution, discussed in Chapter 8; and the decline of an institution that provided contract enforcement in impersonal exchange, discussed in Chapter 10. In the case of the merchant guilds, the related institution fostered the expansion of trade based on rulers' concerns about losing their reputation for protecting the rights of foreign traders. Rulers valued this reputation because they gained from custom duties paid by the traders. Expansion of trade, however, reduced the value of customs paid by the marginal merchant. Initially, the reputation-based institution that motivated rulers to respect the rights of German merchants was based on the threat by the abused merchant and his

[25] The extent to which the political system maps the public's preferences into political outcomes depends on its institutional details.

close associates to cease trading. As trade expanded, however, the reduction to the ruler in the value of the future customs of the marginal merchant undermined the operation of this institution. Additional supporting organizations that rendered credible retaliation by a sufficiently large number of merchants were required.

This historical episode illustrates how the fact that actors play against the rules creates quasi-parameters; changes in various aspects that are incorporated into the rules of the game influence behavior only when those who observe them reveal them through their behavior. At no stage in the process of institutional undermining did merchants directly observe changes in the benefits and costs to rulers of abusing them. Only when a ruler's behavior revealed to the merchants that the institution was no longer self-enforcing did they act to introduce a new institutional element to reinforce the failing institution.

6.7 Concluding Comments

By analyzing reinforcing processes, this chapter examined why and how the behavior induced by self-enforcing institutions influenced their long-term stability. Behavior in equilibrium can gradually alter quasi-parameters in a way that causes institutions to be self-enforcing in a larger or smaller set of situations. Hence institutional equilibria are subject to endogenous change, both indirectly and directly. They do so indirectly by making them more or less sensitive to exogenous shocks. Institutional behaviors directly influence rates of institutional change, for unless a self-enforcing institution is (weakly) reinforced, it will change in the long run. Either the associated behavior will no longer be self-enforcing or new institutional elements will be required to support it.

Endogenous change in this perspective is driven by marginal shifts in the value of quasi-parameters. Such shifts make the institution more or less sensitive to environmental changes, and they can render an institution no longer self-enforcing in a given environment. Analytically, one can combine the study of self-enforcement and reinforcement by first examining an institution's self-enforceability while considering quasi-parameters as fixed and exogenous, then examining the implied reinforcing processes, and finally examining the long-term implications of these processes on the institution's endogenous rate of change.

Several methodological and substantive extensions to these insights into the study of endogenous institutional change are called for. First, the analysis relies on the repeated-game framework, but furthering the analysis of self-reinforcement will benefit from a more explicitly dynamic analytical framework that is only hinted at by the formal model presented here. Second, statistical tests may strengthen the contextually based game-theoretic analysis of institutional change. Unless the observable implications of models of reinforcement are statistically validated over a range of cases outside the set of cases from which the theory was developed, there will remain a tautological residue on those models. However, statistical tests of the observable implications of the model on aspects of the society that were not analyzed in the formation of the model further lend support to the analysis's validity. For example, an observable implication of the model of the two Italian city-states is that over time there would be more interclan exogamy in Venice than in Genoa. Showing that this was the case would help overcome charges of tautology. Furthermore, statistical tests will also allow us to assess the relative importance of endogenous versus exogenous sources of institutional change. Third, the analysis emphasized the importance of quasi-parameters but only began to explore the features of institutions that foster reinforcing or undermining changes in quasi-parameters in various situations.

Substantively, much work remains. The theory presented here concentrates mainly on beliefs (albeit noting the importance of norms in reinforcing institutions); a parallel theory has to be developed regarding norms. The relevant issues are many: Under what conditions is behavior internalized as morally appropriate and hence reinforced? What determines the relative weights in preferences between one's normative behavior and behavior that is materially beneficial? Addressing this question is central to understanding when economic (materialistic) considerations will or will not undermine normatively appropriate but economically unrewarding institutionalized behavior.

More generally, we have no theory to explain the factors that determine the extent and speed of reinforcing processes. What factors, for example, determine the extent of intentional and habitual behavior? What organizations respond to the risks implied by the institutionalized rules rather than to the rules of the game? What determines the ability of individuals to manipulate institutionalized rules?

By introducing and elaborating on the concepts of quasi-parameters and institutional reinforcement, this chapter provides a framework for integrating the study of self-enforcing institutions with that of endogenously induced institutional change. This approach can be extended, as it is in the next chapter, to examine why and how self-enforcing institutions influence the direction of institutional change.

Annex 6: A Model of Institutional Reinforcement

Consider an infinitely repeated prisoners' dilemma game in which the period t = 0,1,... stage game actions and payoffs to the two players are:

1 \ 2	c	d
c	b_t, b_t	$-k$, $b_t + e$
d	$b_t + e$, $-k$	0, 0

where $b_0, k, e > 0$, and players share a common discount factor $\delta \in (0, 1)$. The model has four parameters: δ, b_0, k, and e. b_t, is a quasi-parameter, since it can be affected by the institution in place. The institution we are interested in is the one generating cooperation, that is, stage-game play of (c,c).

Definition: Cooperation has a positive (negative, neutral) reinforcement if play of (c,c) in period t implies $b_{t+1} - b_t > (<,=) 0$. Standard models of repeated prisoners' dilemma take cooperation to have neutral reinforcement. For simplicity, I assume that the change in cooperation payoffs under any reinforcement mechanism is fixed over time.

Assumption: For all t, $b_{t+1} - b_t = \epsilon$ with $\epsilon > (<,=) 0$ under positive (negative, neutral) reinforcement. In what follows, the equilibrium notion is subgame perfect Nash equilibrium. To avoid introducing complicating notation and terminology, the presentation is somewhat informal.

Case 1: Knowledge about Reinforcement: Consider the case in which players are aware of the reinforcement mechanism.

Claim 1: The cooperation institution is self-enforcing over a larger range of discount factors under positive reinforcement than under neutral reinforcement.

Proof: Fix the period as τ. It is easily seen that cooperation can be a self-enforcing institution under neutral reinforcement if and only if

(1) $$\delta \geq \frac{e}{b_\tau + e}$$

Suppose there is positive reinforcement. Recall that $\epsilon \equiv b_{\tau+1} - b_\tau > 0$ under Nash reversion. (Specifically, playing defect every period.) If players follow Nash reversion, then on the equilibrium path their payoffs will be strictly larger than $b_\tau + (b_\tau + \epsilon)\frac{\delta}{1-\delta}$, while deviating yields $b_\tau + e$. Hence cooperation is incentive compatible if $e \leq (b_\tau + \epsilon)\frac{\delta}{1-\delta}$, which can be rewritten as

(2) $$\delta \geq \frac{e}{b_\tau + e + \epsilon}$$

Because $\epsilon > 0$, the right-hand side of expression (2) is strictly smaller than the right-hand side of expression (1), which proves the claim. Q.E.D.

Claim 2: Under negative reinforcement, cooperation is not a self-enforcing institution.

Proof: The proof is straightforward by backward induction, given that payoffs from mutual cooperation decrease by ϵ every period if players have cooperated in previous periods.

The institution of cooperation can thus be self-enforcing only under neutral or positive reinforcement. Under positive reinforcement, the institution is positively reinforcing, because the right-hand side of expression (2) decreases over time (because b_t increases), causing the equilibrium to hold for a larger range of δ over time. By similar reasoning, the institution is neither positively nor negatively reinforcing under neutral reinforcement, because the range of δ over which it is self-enforcing is identical in any period t.

Case 2: Ignorance about Reinforcement. Now consider the case in which players are unaware of the reinforcement mechanism. In each period players observe b_t and imagine that this value remains fixed in all future periods regardless of their actions. If cooperation can be supported in equilibrium, it can be done with Nash reversion. In any period τ, this is incentive-compatible if and only if $b_\tau + e \leq \frac{b_\tau}{1-\delta}$, or equivalently, if and only if

(3) $$\delta \geq \frac{e}{b_\tau + e}$$

The right-hand side of expression (3) is strictly decreasing in b_τ. Hence if cooperation produces positive reinforcement, the range of δ for which Nash revision is self-enforcing increases over time (i.e., the institution is positively self-reinforcing). If the institution is self-enforcing in some period τ, it will be self-enforcing in all periods thereafter.

If cooperation produces negative reinforcement, the institution is negatively self-reinforcing. Indeed, with negative reinforcement, for any δ and any starting value b_0, there is some (possibly large) t such that cooperation is no longer self-enforcing at period t. At t the institution changes to defect, defect. The gradual erosion of the gains from cooperation implies that at some point, the future gains from cooperation are smaller than the present gain from defecting.