A Theory of Endogenous Institutional Change

AVNER GREIF and DAVID D. LAITIN *Stanford University*

This paper asks (a) why and how institutions change, (b) how an institution persists in a changing environment, and (c) how processes that it unleashes lead to its own demise. The paper shows that the game-theoretic notion of self-enforcing equilibrium and the historical institutionalist focus on process are both inadequate to answer these questions. Building on a game-theoretic foundation, but responding to the critique of it by historical institutionalists, the paper introduces the concepts of quasi-parameters and self reinforcement. With these concepts, and building on repeated game theory, a dynamic approach to institutions is offered, one that can account for endogenous change (and stability) of institutions. Contextual accounts of formal governing institutions in early modern Europe and the informal institution of cleavage structure in the contemporary world provide illustrations of the approach.

Why and how do institutions change? How do institutions persist in a changing environment? and How do processes that they unleash lead to their own demise? These questions are particularly difficult to address when institutions are viewed from a game-theoretic perspective in which they are considered self-enforcing, and in which all behavior is generated endogenously.[1] In reviewing the merit of various approaches for studying political institutions, Hall and Taylor (1996) noted the advantage of the game-theoretic perspective. It demands a "precise conception of the relationship between institutions and behavior" (p. 950) and thereby allows us to explain "why existing institutions continue to exist" (p. 952).

A challenge that this line of research faces, however, is the difficulty of addressing the issue of how institutions change endogenously. After all, a self-enforcing institution is one in which each player's behavior is a best response. The inescapable conclusion is that changes in self-enforcing institutions must have an exogenous origin. No one has an incentive to deviate from the behavior associated with the institution. As noted by Hall and Taylor (1996, 953), "The 'equilibrium' character of the rational choice approach to institutions embroils such analysts 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." Endogenous institutional change appears, then, to be a contradiction in terms.[2]

Indeed, the analysis of institutional change from this perspective has mainly concentrated on the dynamics following environmental changes—that is, changes in parameters exogenous to the institutions under study.

Development of the game-theoretic perspective on institutions thus requires extending it to studying institutions as a product of an historical process in which institutions endogenously change. Furthermore, doing so will enhance integration of this perspective with complementary perspectives, such as that of institutional path dependence in economics (David 1994; Greif 1994; North 1990) and historical institutionalism in political science (Hall and Taylor 1996; Pierson and Skocpol 2002; Thelen 1999). To integrate the game-theoretic perspective on institutions with complementary perspectives requires a more dynamic approach than presently offered by the notion of self-enforcing institutions. We suggest that in developing such an approach it is imperative to introduce two related concepts into the institutions-as-equilibria research program: *quasi-parameters* and *institutional reinforcement*.

Before we address dynamics, we can foreshadow our contribution by noting the distinction between parameters and variables in the institutions-as-equilibria perspective. Parameters are exogenous to the institution under consideration. If parameters change, therefore, there is a need to study the implied new equilibrium set and, hence, the new possible institutions. Variables, on the other hand, are determined endogenously by the institution under consideration. Institutional analysis from the institution-as-equilibria perspective thus typically concentrates on a single transaction—e.g., a ruler securing property rights for the ruled—and examines as variables possible self-enforcing behavior in it—e.g., security of property rights in this transaction—for a given set of parameters.

In contrast, we hold that it is conceptually sound and analytically tractable to recognize that some aspects

Avner Greif is Professor of Economics, Mail Code 6072, Stanford University, Stanford CA 94305 (avner@stanford.edu).

David D. Laitin is Professor of Political Science, Mail Code 6044, Stanford University, Stanford, CA 94305 (dlaitin@stanford.edu).

This paper is based on Chapter 6 of Greif (n.d.), which was presented at the Annual Meeting of the ISNIE (Berkeley), September 2001. The authors acknowledge the help received from Avanish Dixit, Henry Farrell, James Fearon, Judith Goldstein, Peter Gourevitch, Leonard Hochberg, Peter Katzenstein, Timur Kuran, Bentley Macleod, Douglass C. North, Daniel Posner, Rudolf Richter, Shankar Satynath, Kenneth Shepsle, Kathleen Thelen, and Barry Weingast in their comments on early versions of the manuscript.

[1] Studying institutions using game theory was suggested by Schotter 1981 and Ullmann-Margalit 1977. For some recent advances, see Aoki 2001; Calvert 1995; Greif 1993, 1994, 1997, 1998, n.d.; Moriguchi 1998; and Weingast 1996.

[2] Although we take this criticism of the game-theoretic contribution to be fundamentally fair, it should be pointed out that *ex ante* creation of institutional arrangements can be predicated on variables whose realizations do not occur until *ex post*. Once such a realization occurs, the institution can change as part of a dynamic equilibrium. See Muthoo and Shepsle 2003 for an example. In our discussion of stability in the face of parametric shift, however, we argue that it is appropriate and realistic to model institutions when the long-term implications of a shift in variables are not *ex ante* foreseen.

of the situation should be considered as parametric in studying self-enforceability but as variables in studying institutional dynamics. It is appropriate to inquire whether the institution—which we analyze as a game-theoretic equilibrium—endogenously affects aspects of the situation apart from behavior in the transaction under consideration. Such aspects should be considered as parametric in studying self-enforceability but as endogenously determined—and thus variable—in the long run. Those parameters that are endogenously changed in this manner and with this effect are *quasi-parameters*. Analysis needs to recognize that 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 rendering a particular behavior an equilibrium. Yet the distinction among a parameter, a variable, and a quasi-parameter is not rigid, and is based on empirical observations. If self-enforcing outcomes affect the values of one or more parameters supporting the observed equilibrium but in a manner that would only lead 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 would find it best to adhere to the behavior associated with it. When self-reinforcing, exogenous changes in the underlying situation that otherwise would have led an institution to change would 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. Hence, institutions can be self-undermining and the behaviors that they entail can cultivate the seeds of their own demise. However, institutional change will endogenously occur only when the self-undermining process reaches a critical level such that past patterns of behavior are no longer self-enforcing.

Because of our interest in change, this perspective builds on the observation stressed in historical institutionalism regarding the importance of historical processes. As noted by Pierson and Skocpol (2002, 698), studying institutions "usually means to analyze processes over a substantial stretch of years, maybe even many decades or centuries." And by bridging the game-theoretic and historical perspective—by examining the relationships among 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—we enrich both. While extending the game-theoretical perspective, we also contribute toward mitigating what Hall and Taylor (1998), among others, identify as an important obstacle for furthering historical institutionalism. "[Important] junctures," they reason, "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 (266). What we suggest here is a way of identifying and studying how the new conditions that an existing institution entails can overwhelm the self-enforcing behavior associated with it.

Our approach speaks to a central concern in political science today—*viz*., how to explain both institutional stability and change. We exploit the notion of institutions as game-theoretic equilibria without excluding the possibility that institutional change may be endogenous to the character of institutions. We do so by introducing two conceptual innovations—quasi-parameters and institutional reinforcement. In this paper we develop these concepts theoretically and apply them empirically.

THEORETIC FOUNDATIONS

Game Theory

Classical game theory has provided a conceptual apparatus for the analysis of self-enforcing institutions, central to which are shared beliefs. The recognition that such beliefs are essential to institutions goes back to Durkheim ([1895] 1950), who viewed institutions as being composed of beliefs and modes of conduct shared in a collectivity. Game theory advances this line of analysis: It provides an explicit analytical framework enabling us to deductively restrict the set of (rational) shared beliefs capturing individuals' expectations with respect to actions that others will take in various contingencies. As noted by Greif (1994), since shared beliefs are identical and commonly known, when players play their best response to these beliefs, the set of permissible beliefs is restricted to those that are self-enforcing. Hence this specific subset of beliefs can be formalized as a set of probability distributions over an equilibrium strategy combination. Each probability distribution reflects the expectation of a player with respect to the actions that would be taken on and off the path of play. In equilibrium, only shared beliefs corresponding to self-enforcing behavior can rationally prevail (Calvert 1995; Greif 1993, 1994).

Analytically, then, after specifying the game, game-theoretic analysis proceeds to restrict deductively the set of admissible institutions to those that are self-enforcing by examining the equilibrium set. An important contribution of this framework is making explicit the dependency of possible equilibria on the parameters—such as the payoffs from various actions, time discount factors, risk preferences, wealth, and the number of players—of the underlying game. In particular, the framework highlights the conditions under which an exogenous change in parameters will undermine institutional self-enforcement. As summarized by Weingast (1996, 180), in this framework "institutions are the endogenous variable, adjusting as exogenous circumstances change."

Pushing this perspective to its logical conclusion, is it not possible to study such self-enforcing institutions through the specification of a game devoid of any human constructions? As articulated by Calvert (1995, 59), such an attempt entails considering the game as "just a description of underlying physical realities: if people behave in a certain combination of ways, nature responds with certain goods or conditions. Any additional structure 'instituted by the collectivity' must be described as the behavior patterns of individuals and their expectations about the behavior of others." As a matter of practice, however, we always have to take some human constructs as given, as parametric to the analysis. The existence of language and money is assumed in most applications. Organizations such as communities, firms, parties, and legislating organizations are often assumed as well. We often take them as parametric as a matter of convenience, although such socially constructed features (but not all, at least in a tractable model) can also be examined from the self-enforcing institutions perspective (Greif 1994, 943).

Despite recent advances, we do not claim that game theory is sufficient for institutional analysis (Greif, n.d., part III). We recognize that while game theory provides a useful analytical tool for studying self-enforcing beliefs and behavior in a given situation, by virtue of its sparseness, it does not capture fully the complexity of the interrelationships between individuals and the institutions influencing their behavior. Accordingly, we define an institution not as an equilibrium but in a way that distinguishes between the object of study—institutions—and the analytical tools used to study them. We define institutions as a system of human-made, nonphysical elements—norms, beliefs, organizations, and rules—exogenous to each individual whose behavior it influences that generates behavioral regularities.[3]

Some unpacking is in order. Institutionalized norms and shared beliefs provide motivation. Organizations are institutional elements that influence the set of beliefs and norms that can be self-enforcing in the transaction under consideration. Rules are behavioral instructions that facilitate individuals with the cognitive task of choosing behavior by defining the situation and coordinating behavior. When we study self-enforcing institutions we thus consider them as composed of self-enforcing institutional elements that motivate, coordinate, and enable individuals to follow particular regularities of behavior.

Historical Institutionalism

Historical institutionalists focus on process rather than equilibria (Thelen 1999) and have criticized game theory's concentration on self-enforcing beliefs because of its bias toward institutional stability. In this school, it has been suggested that the ideal way to study process and stability is through the analysis of positive and negative feedback loops (Ikenberry 1994). Along these lines, Levy (1999) has shown that to the Tocquevillians' chagrin, as the French state took greater control in directing the national economy, it sidestepped and eventually undermined local business associations. This had dire consequences. In the 1980s the left attempted a bottom-up approach to economic regeneration. This failed because the state could find no "collective business partner" to energize development of small and medium enterprises in localities. Decentralization requiring an active civil society could no longer be a policy option for the French state because in an earlier era civil society was decimated. Levy dubs this "Tocqueville's revenge." Here we see an example of a negative feedback loop that undermined civil society such that the reform-minded socialists could not rely on local initiative.

Negative feedback may lead to and be part of critical junctures in which new institutions are created (Collier and Collier 1991; Katznelson 1997; Thelen 1999). Ertman (1997), for example, argues that the context in which states enter the competitive international system is a critical juncture, undermining old institutional forms and creating new paths toward patrimonial or bureaucratic states. Although not as attuned to process as advocates of historical institutionalism have called for, studies of critical junctures illustrate the sources of new institutions in societies whose previous institutions, due to path dependence, had been resistant to change.

If game theorists have difficulties accounting for change, historical institutionalists writing about critical junctures have faced the problem of too easily accounting for change, making institutions seem "fluid" (Thelen 1999, 397). Thelen further argues that this tradition does not adequately analyze how the reproduction of these institutions occurs. Many assume that the institutional innovation "filled the political space" or "crystallized" in ways that are difficult to alter. But she points out that the world is littered with organizations formed in critical junctures that could not survive changing environments. Thelen suggests that feedback loops should be able to account for both stability and change through a focus on "the particular mechanisms of reproduction" (397). To illustrate her point, she provides a vignette of two post–World War II party systems, in Sweden and Italy. Both were headed by a hegemonic party. Sweden's party was sustained through programmatic successes; Italy's by patronage. The mechanisms of reproduction explain not only stability, Thelen suggests, but the modes of change. When Sweden's once hegemonic party was defeated in the polls, its history of programmatic success allowed for new recruitment and the party thereby survived in new competitive circumstances. But when Italy's was so defeated, and the scandals associated with its recruitment through patronage became clear, it virtually collapsed in ignominy, reconfiguring the political

[3] Greif n.d., chaps. 2, 5. Cf. Calvert 1995. Where Calvert defines institutions as an equilibrium (with a single parameter, beliefs), we identify institutions based on whether the set of elements (some parametric, others quasi-parametric) induces regularities of behavior while being exogenous to each individual whose behavior they influence. This allows us to assess the equilibrium attributes of an institution without assuming that by definition the institution is in equilibrium.

landscape. Only by understanding the mechanisms of reproduction, Thelen concludes, can you understand the processes of change. So analysis of stability and change, she insists, should not be different analytic fields, but are inseparable. We agree.

Pierson's (2000) application of increasing returns (equivalent for him to positive feedback) is historical institutionalist in mode of argument, yet, in the hope of bridging the divide with game theory, it emphasizes stability. He argues that given the nature of politics (its opacity, its fundamental reliance on collective action, its high density of institutions, and its granting authority to actors who want to enhance power asymmetries), outcomes remain stable because "the costs of switching from one alternative to another will . . . increase markedly over time" (251). His earlier study on how and why President Reagan and Prime Minister Thatcher had such difficulty retrenching from the welfare state is a model example of how constituencies that have conditioned their actions on these programs grew over time, thereby making them more robust against political challenge (Pierson 1994).

Like Pierson, we seek to bridge the divide between game-theorists and historical institutionalists. Yet we are not fully satisfied with Pierson's conceptual apparatus. First, Pierson relies on a vague specification of parameters. In his descriptions of increasing returns processes, there is no indication of which parameters are changing in the existing equilibrium to make it more robust against alternatives. We argue that increasing returns (to the beneficiaries of a program, for example) do not necessarily translate to supporting that program. What we need to have is an articulation of why the programs were an equilibrium in the first place. With such an articulation, we could highlight the particular factors that would make a welfare program an equilibrium for a larger or smaller parameter set. In other words, he needs better to delineate the mechanisms by which increasing returns take place. We delineate this mechanism through analysis of quasi-parameters.

Second, Pierson (2000) treats negative feedback and increasing returns as if they were distinct—the former is exogenous to political processes and the latter endogenous (2000, 265–66). But surely all institutions endogenously unleash processes of stability and change simultaneously, implying that we need a unified framework that can analyze both processes. It could well be, as Pierson (2000, 253) recognizes, that strong positive feedback produces stability in the short term, but slowly accumulating negative feedback can at the same time induce institutional collapse. An approach to institutions is incomplete, however, when it focuses on the short term, leaving complementary processes with long-term implications to the background. Analyzing reinforcement and undermining within the same framework give advantage to our approach.

Third, Pierson's approach focuses primarily on payoffs and misses other parametric shifts having consequences for institutional stability. Consider the case of social security. One of the parameters in any analysis of social security as an institution is demography. A successful social security policy will tend to increase the number of the elderly (citizens will live longer if they get income support) and to reduce the number of children (citizens will invest in fewer children if they are assured of old age support from the state). Other things equal, with more elderly and fewer young, the expected payoffs from social security will decline. Although this reflects decreasing returns, support for the institution would not decline. As the ratio of old to young increases, the political support for social security will increase. Our notion of reinforcement enables us to capture such situations as it focuses on the general extent to which the institution is self-enforcing rather than only on the benefits it entails.

In order to overcome these problems, but in the spirit of Pierson's contribution, we develop a conceptual framework that takes into account the processes he identifies but is more specific about the elements of those processes. We return to historical institutionalist findings in our conclusion to show how they might benefit from being incorporated in a unified game-theoretic framework.

A NEW APPROACH TO ENDOGENOUS INSTITUTIONAL CHANGE

Game-theoretic analyses of institutions have traditionally focused on studying the relationships between the rules of the game and how regularities of behavior—cooperation, wars, political mobilization, social unrest—are generated in the particular transaction under consideration. Repeated game theory turned out to be particularly useful in exploring the relationships between the details of the transaction (the relevant parameters captured in the rules of the stage game) and the set of possible self-enforcing beliefs and behaviors. In particular, the analysis allows us to see whether a particular strategy combination—a plan of behavior—is a subgame perfect equilibrium in which all threats and promises are credible.

While the fruits of this analytical posture are many, and its focus on regularities of behavior in a particular transaction is useful, it diverted attention away from considering an institution's other possible ramifications that go beyond the behavior it implies in the transaction under consideration. 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. Even if not 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-theoretical framework, such influence implies a dynamic adjustment of *variables* that, if this influence had been ignored, would have been considered *parameters* in the stage game.

Stability in the Face of Parametric Shift

While the folk theorem exemplifies the game-theoretic insight that a multiplicity of self-enforcing institutions

is generically associated with a particular parameter set, the theorem also highlights a corollary to this insight: A particular equilibrium can generically be sustained over a broad range of parameters. If a strategy combination is an equilibrium, it would generically be an equilibrium in some parameter *set*. As long as the actual parameters are in this set, game theory does not predict that the associated beliefs and behavior would not prevail. Game theorists accordingly have long recognized that game theory does not *predict* behavioral change following a parametric change.

Indeed, as Schelling's (1960) seminal work on focal points reminds us, there are good reasons that individuals would continue to follow past patterns of behavior even under conditions of marginal parametric change. This is the case for at least three interrelated reasons, knowledge, attention, and coordination. Elaborating on these reasons requires considering in more depth the role of institutionalized—socially articulated and distributed—rules of behavior.

Knowledge. Why are institutionalized rules such a salient feature of societies? Why, for example, doesn't each member of a society develop a distinctive rule of behavior for him- or herself through experimentation and induction? In addressing these questions, it is useful to first note the contributions of game theory to our understanding of the complexity of decision-making in strategic situations and hence, indirectly, the role of rules. Game-theoretic models usually assume that players have a complete and closed model and correct common priors. Each player has complete information about the details of the situation, including others' preferences, the magnitude of various parameters, and various causal links. When such information is missing, the players assign the correct prior probabilities to all possible values of the unknown parameters. All players assume that their opponents model the game exactly as they do and they too assign the same correct probability to all unknown parameters.

These assumptions do not capture the world as we know it. Indeed, the complete model is demanding. Interactions among individuals are carried out in a complex environment containing many unobserved features that affect one's decision about how to act. These features include others' preferences and the specific magnitude of various parameters, such as wealth, the time discount factor, and outside opportunities. The computational complexity required to solve even a moderately complicated game is daunting.

Bounded rationality learning models is the game-theoretic response to its own unrealistic assumptions (e.g., Rubinstein 1998). Such investigations have been conducted, by and large, while retaining the assumption (that also prevails in classical game theory) that individuals are not guided by institutionalized, social rules. Learning is done in an individualistic and atomistic manner. Modeling such learning processes, to make them tractable, has involved replacing the assumptions of classical game theory with a new set of questionable assumptions, the central one being that individuals are myopic (Fudenberg and Levin 1998; Marimon 1997; Young 1998). Myopic behavior can imply, however, such unreasonable behavior as not performing a costly experiment no matter how high the resulting expected return might be. When such restrictions were not imposed, however, the resulting analysis became too complicated to provide a convincing account of how individuals in general learn.

In contrast, models that allow for socially transmitted rules are able to substantiate that learning can lead to regularities of behavior—to an equilibrium—without having to impose the restrictive conditions of either classical game theory or bounded rationality learning models. Kalai and Lehrer (1993, 1995) considered learning in a repeated game in which individuals share a cognitive system but all know only their own payoff matrix and discount factor. The players have the same cognitive understanding of the game but no player knows all the relevant parameters of the model. All players, observing the outcomes of the game, can develop only a subjective evaluation of the parameters and others' strategies. This analysis assumes that individuals are subjectively rational and does not impose the restriction that players assume that others are rational. Each one starts with subjective beliefs about the strategies used by each of their opponents and uses these beliefs to compute one's own optimal strategy. As the game unfolds and the players update their beliefs, players develop their own subjective models of the situation.

Analyzing this process of learning, Kalai and Lehrer show how an initial "grain of truth" regarding others' behavior is sufficient for individuals to learn independently how others will play. Specifically, if each player's initial subjective beliefs assign a positive probability to the events that would occur in the play of the game, then learning will lead each player, over time, to be able to appropriately predict the behavior of the others. If this is the case, these individuals will converge in finite time to play a Nash equilibrium in the real game.[4] Subjectively developed beliefs thus converge on equilibrium beliefs. Equilibrium beliefs are reached not by directly observing the relevant parameters of the game but rather by convergence aided by institutionalized rules. Socially articulated and distributed rules provide individuals with the initial "grain of truth" to develop subjective beliefs regarding others' behavior. Institutionalized rules assist individuals in forming beliefs—in placing a probability estimate—about what others will do. But these rules do not have to be accepted as objectively correct. As long as subjectively rational individuals accept the behavior associated with the rule as possibly correct, and this rule assigns positive (initial) probability to the Nash behavior, the processes of learning leading to a Nash equilibrium transpire.

While the Kalai and Lehrer argument is intuitive, its technical analysis rests on the assumption that individuals will use Bayesian updating, which is not necessarily how people update their beliefs. But if people do not employ the logic of probability theory that Bayesian

[4] The equilibrium satisfies the Nash or the epsilon-Nash restrictions. These details are unimportant here.

updating captures, institutionalized rules specifying others' behavior are arguably even more important in leading to regularities of behavior. Indeed, a sufficient condition for a Nash equilibrium is that individuals, who know only their own action sets and preferences, have an accurate prediction of what others will do. Knowledge (or common knowledge) of the rules of the game is not necessary (Aumann and Brandenburger 1995).

Institutionalized rules only assist in forming, but do not determine, beliefs because retrospective individuals compare outcomes with expectations. Only rules in which expected and actual behaviors correspond to each other will lead to the reproduction of these beliefs as they are confirmed by observed behavior. Hence, it is appropriate to restrict social rules that can establish themselves in a society and be followed—that can become institutionalized—to those that correspond to an equilibrium.

Because institutionalized rules constitute an equilibrium in the responses of individuals to these rules, they not only assist individuals in choosing behavior, but also aggregate, in equilibrium, the dispersed information that each of these individuals has. In other words, these rules both provide individuals with the information they need to make decisions regarding how to act as well as aggregating the information privately held by each of them. Institutionalized rules of behavior aggregate information in a compressed form and direct individuals to play an equilibrium strategy in the underlying game, although the assumptions held in classical game theory do not hold. And because such rules correspond to an equilibrium, individuals are seemingly rule-followers; they follow the rules associated with the social position they occupy.

Just as market prices aggregate the economic agents' private information, game theory reveals similar relationships between institutionalized rules of behavior and each individual's private information. These rules inform each individual about others' expected behavior. But the only rules that can correspond to actual behavior are those in which all individuals, basing their decisions on private information, find it optimal to follow. In an institution, institutionalized rules aggregate the private information of all agents, providing each with sufficient statistics to make an informed decision.[5]

In our analysis of transactions in which individuals are guided by institutionalized rules, it is therefore appropriate to assume that players share common knowledge of the rules of the game. The information compressed in socially transmitted rules enables individuals without knowledge of all the relevant parameters and causal mechanisms, and with limited computational ability, to act in a manner that leads to equilibrium behavior. Once equilibrium is achieved, because individuals do not observe relevant parameters and lack full comprehension of causal relationships, the best they can do is to perceive the world as stationary as long as observations—including those conveyed through others' behavior—do not contradict this perception.

Hence, past behavior would reign despite marginal parametric changes because institutionalized rules enable individuals with limited knowledge and information to choose behavior. And thus, behavioral rules learned in the past are the best predictor of future behavior. As long as others' behaviors (that one neither observes nor understands their causal underpinning) do not reflect a change in the parameters, one would not change one's own behavior either. Conversely, when the parameters that one observes marginally change, one faces the problem of which behavior to follow in the new situation, given the multiplicity of self-enforcing behaviors. Expecting others to continue following the previous institutionalized rules of behavior, it is rational to continue following the past patterns of self-enforcing behavior.

Attention. Past patterns persist also because what one sees, knows, and understands in a given situation also reflects the amount of attention one devotes to the task. Attention is a scarce resource. Institutionalized rules come to the rescue. They enable one to choose behavior in complicated situations while devoting limited attention to decision-making in noninstitutionalized situations.[6] Parametric shifts that can be noticed if more attention would have been devoted to observing them may go unnoticed, further contributing to lack of behavioral change in response to marginal parametric changes. Moreover, those who observe the parametric shift and can bring it to the attention of others may not have the incentive to do so. And an observed marginal parametric shift is not likely to induce decision makers to devote the cognitive resources required for considering whether to change their behavior or not. People do not stop at every choice they make in their lives and consider their optimal response (DiMaggio and Powell 1991).

Coordination. Coordination failure is the third reason why a marginal parameter shift does not necessarily lead to changed behavior. When a situation marginally changes, individuals face the problem of which behavior to follow in the new situation, given the multiplicity of self-enforcing behaviors. Because people do not share the expectations that some new equilibrium behavior will be followed, they are likely to rely on past rules of behavior to guide them and to continue following past patterns of self-enforcing behavior. This would be the case even when there are individuals and organizations with the ability to coordinate on new behavior. There are many reasons why 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.

[5] This discussion counters the common assertion that game theory is inappropriate for studying institutions because it assumes that the rules of the game are common knowledge. Furthermore, this assertion is misleading to begin with because common knowledge is neither necessary nor sufficient for the Nash equilibrium condition to hold.

[6] Simon (1976) has argued that habits similarly serve the function of directing attention to selected aspects of a situation.

Quasi-Parameters and Reinforcement

Many features that are usually taken as parameters in the repeated game formulation share two properties: First, they can gradually be altered by the implications of the institution under study, and second, their marginal change will not necessarily cause the behavior associated with that institution to change. They do not cause the behavior associated with that institution to change because the changes of these features and their ramifications on the institution are not *ex ante* recognized, anticipated, directly observed, appropriately understood, or payed attention to. These features are neither parameters (as they are endogenously changed) nor variables (as they do not directly condition behavior); they are quasi-parameters. Because changes in quasi-parameters and their implications are not recognized by the actors, we have to consider them as parametric—exogenous and fixed—in studying the self-enforcing property of an institution in the *short run*, but we have to consider them as endogenous and variable when studying the same institutions in the *long run*.

Changes in quasi-parameters that an institution implies can reinforce or undermine it. 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—other parameters—than would otherwise have been the case. A self-enforcing institution that reinforces itself is a *self-reinforcing* institution. But a self-enforcing institution can also undermine itself when the changes in the quasi-parameters that 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. An institutional change is a change in beliefs, and 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.[7] 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, it will eventually reach a situation in which the behavior associated with it is no longer self-enforcing. Endogenous institutional change would follow.

Considering reinforcement, however, highlights the importance of another, indirect way that an institution endogenously influences its change—when it influences the magnitude and nature of the exogenous shocks that will be necessary to cause the beliefs and behavior associated with that 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 the previous one. The behavior associated with it would be self-enforcing even in situations in which, previously, this would not have been the case. The opposite holds in the case of an institution that undermines itself. An institution, by reinforcing or undermining itself, indirectly influences its rate of change by determining the size of an external change in parameters required to render behavior associated with it to cease being self-enforcing.

Institutions can change due to endogenous processes, exogenous shocks, and combinations of both. The exact 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 quasi-parameters are observable and their importance well understood, decision-makers might actually realize that past behavior is no longer self-enforcing and the mechanism directly leading to institutional change will be intentional. Intentional selection of alternative behaviors, specification of new rules through collective decision-making, and intentional introduction of organizations are common manifestations of this mechanism. But an institution can cease to be self-enforcing due to changes in quasi-parameters that are unobservable, uncertain, and unrecognizable. In such cases, 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 reveal a new institutional equilibrium.[8] In either case, learning is slow and institutional change is rare. It may take a long time for self-undermining to be reflected in new behavior.

Hence, institutional change should have a quality of punctuated equilibria (Krasner 1984), where change is in actuality evolutionary but apparently abrupt, typically associated with a "crisis" revealing that the previous behavior is no longer an equilibrium. But because of the cognitive, coordinative, and informational content of institutionalized rules and the nature of other institutional elements such as beliefs and norms as properties of individuals, responses to observed institutional failure would not constitute a complete departure from the past. On the contrary, these responses would often

[7] Here we focus only on the issue of endogenous institutional change due to self-reinforcement and undermining, but the above observations regarding the nature of institutions, institutionalized rules, and beliefs enable us to address related issues—e.g., 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.

[8] Game theory indicates the importance of uncertainty in these processes. If the eventual collapse of the institution is known and anticipated to prevail at a particular point in 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 set of behaviors that can prevail in an infinitely repeated game.

constitute what can be referred to as "institutional refinement." New institutions organically evolve (or are intentionally designed) through changing, introducing, or manipulating institutional elements while supplementing existing elements (or responding to their failure to generate desired behavior). This is the foundation for a dynamic theory of institutions that, unlike many strands in historical institutionalism, does not overpredict change (Greif n.d., chap. 9).

SELF-REINFORCEMENT—TWO PAIRED COMPARISONS

In this section, we illustrate our dynamic approach to institutional change through two paired comparisons of institutions—political regime in Venice and Genoa and cleavage structure in Nigeria and Estonia. A capsule summary of our framework is given in Table 1. One pair (Venice and Genoa) is based on historical archives where parametric change is part of the long durée; the other (Nigeria and Estonia) is based on fieldwork where parametric change is a projection of trends. In both comparisons, each of the institutions is self-enforcing, but only one of them is self-reinforcing. In the Genoa/Venice comparison, we show why Venice's early modern political institutions were self-reinforcing but Genoa's were not, leading Genoa to civil war and economic stagnation. In the Nigeria/Estonia comparison, we show why Nigeria's cleavage structure but not Estonia's is self-reinforcing, leading Estonia but not Nigeria to potential changes in the political organization of interests. In both comparisons, we highlight the implications of the change in the value of quasi-parameters for self-enforcing institutions.

Our paired comparisons illustrate an as yet implicit aspect of our definition of institutions as systems of elements—organizations, rules, norms, and beliefs. Within any organization, or around any set of rules, there are subsets of coordinated elements that are themselves institutions. Institutions can be identified therefore at different levels of aggregation. In our first paired comparison, we examine an encompassing political institution—the political regime. In our second paired comparison, we take a subset of organizations, rules, norms, and beliefs from the political regime—that of the cleavage structure—and analyze it as an institution as well. Although we cannot examine endogenous shifts affecting all institutional elements—most notably we do not explore empirically changes in rules—our examples below illustrate the usefulness of repeated, complete information models to facilitate the analysis of institutions and their dynamics.

A Tale of Two Cities

Our first comparison is that of late medieval Venice and Genoa.[9] We analyze the *political regime* as an institution. It is a system of elements—the organization of the governing structures, the rules for choosing leadership positions and behavior, the norms of fair distribution of authority and resources, and the rules and beliefs shared by citizens about what fellow citizens would do in unforeseen circumstances.

Historical Background. The residents of the Venetian lagoon established Venice as a political unit in 697 and residents of Genoa organized themselves into a Commune around 1096. These cities became the two most commercially successful Italian maritime city-states.[10] The rise of both cities reflects opportunities for commercial expansion made possible by the relative naval and military decline of Muslim and Byzantine forces around the Mediterranean, particularly during the eleventh century. The residents of Genoa and Venice at the time found themselves in a political vacuum, as both the Byzantine Empire (that claimed sovereignty over Venice) and the Holy Roman Empire centered in Germany (that claimed sovereignty over Genoa) were not in a position to interfere in local political developments.

In both cities, clans and families became the prominent unit of social organization due to the decline of central authority (Hughes 1978). Given this decline, Herlihy (1969, 178) 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."[11] Genoa and Venice were established by an agreement among the strongest clans in each city to cooperate politically for the advancement of their economic interests.[12]

The resulting political institutions governed a particular transaction: motivating individuals—members of the cities' strong clans and families—to relinquish decision-making power and resources in return for political order and the economic benefits of collective action. The political organizations of Genoa and Venice were seemingly identical. Both cities were governed by an oligarchy that, by and large, selected their political leaders, and these leaders were subject to the law. At

[9] For a general discussion of Venice and Genoa histories, see Lane (1973) and Epstein (1996), respectively. The analysis here builds mainly on Greif 1995, 1998, n.d.. See also González de Lara 2004.

[10] Our line of analysis departs from a long tradition in the study of this historical episode. Lopez (1976), in his seminal work on commercial development during this period, did not examine the relationship between political organization and commercial success, maintaining that "the Italian communes were essentially governments of the merchants, by the merchants, for the merchants—an ideal platform for" commercial expansion (71).

[11] See also Herlihy's discussion on pages 174–75, in which he contrasts his analysis with the traditional one (namely, that the history of the family is a history of progressive nuclearization). Even the traditional approach does not dispute the importance of the clan in the late medieval period, but traces its origin to an earlier period.

[12] An agreement for interclan cooperation, however, does not imply that clans were unwilling to use force against each other to advance their particular 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 1998; Tabacco 1989). Genoa's two dominant viscount clans were a product of the feudal world of the time in which one's 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.

TABLE 1. Summary of Paired Comparisons

Case	Theoretical Concept Applied: Institution (Transaction Governed by the Institution)	Shared Institutional Elements	Differentiating Institutional Elements	Quasi-Parameters	Why Self-Enforcing?	Why (Not) Self-Reinforcing?
Genoa	Political regime! (autonomy for order, benefit of collective action)	Organization: clans as fundamental constituents	Podesteria Beliefs: clans are opportunistic Norms: use of force to achieve political goals	Wealth, patronage, identity, strength of popoli and clans, norms of revenge	Mutual deterrence of clans while distribution depends on clans' relative strength	Trade → wealth → patronage strengthened popoli → stronger clan military ability, identity, and norms of revenge → undermining of mutual deterrence
Venice			Magistracy Beliefs: clans would use force to protect the system		Mutual deterrence in the context of even distribution to all clans	Trade → more wealth to distribute → stronger incentive to protect the system → internal peace → Venetian identity → reinforced support for magistracy
Nigeria	Cleavage structure (political support or votes for group favors)	Organization: tribal/national groups Beliefs: everyday primordialism	Low social status of plurality group (Hausas)	ELF	Path-dependent beliefs and organizations of colonial era	Low social status of Hausas and federalism (in oil economy) → increased resources to those who make ethnic/linguistic claims → reinforced cleavage structure
Estonia			High social status of majority group (Estonians)			High social status of Estonians and Unitarism → young Russians assimilate → changing value of ELF → decreasing the realms in which ethnicity is relevant for action

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 able to support interclan cooperation that initially fostered commercial expansion and political order. Yet the subsequent history of the two cities differs. Venice was able to maintain political order in a changing economic environment and to mobilize resources maintaining its economic prosperity even following the decline of its trade with the Far East. Meanwhile, its members' social attachments to the clan structure seem to have slowly declined. Genoa's experience was the opposite. Political order often broke down contributing to its economic decline. Meanwhile clans became more important social and political units over time.

How can these different histories be accounted for despite the similarity in initial conditions, outside opportunities and the basic political structures of the two cities? Both Genoa and Venice had initially developed political regimes that were sufficiently self-enforcing to sustain economic prosperity. Yet Genoa's institutions were self-undermining while Venice's were self-reinforcing. In developing this argument we consider quasi-parameters such as the wealth of the cities, the strength of the popoli, and the social identities of the clans. Understanding these cities' subsequent histories requires considering how these quasi-parameters changed their values as a result of the implications of the institutional equilibria. Changes in the quasi-parameters in Genoa had the effect of undermining political order, making its institutions sensitive to relatively small exogenous shifts in clans' strength, trading opportunities, and level of external threat. The opposite changes, however, transpired in Venice, whose magistracy was self-reinforcing.

To understand these histories and their long-term implications, we examine more closely these cities' institutions. Although the origins of these two distinct institutions are not the focus of the discussion, they probably reflect the institutional heritage of the Doge, a less unequal initial distribution of interclan military might and wealth, and a series of able leaders who coordinated on and developed elements of Venice's institutions.

Genoa. Initially (from 1096 to 1194), Genoa had elected consuls who functioned as the city's political, administrative, and military leaders.[13] These consuls were representatives of the main Genoese clans (Hughes 1978, 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 were guided by the *belief* that clans would challenge each other 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 clan to use its military might to gain political and economic dominance over the city but each clan was deterred from doing so because of the other's military strength. Hence each of Genoa's main clans was motivated to mobilize its resources for interclan cooperation to advance Genoa's economy, but only to the extent to which its ability to deter other clans from militarily challenging it was not undermined.

Early in this period, the relatively high gains from the joint mobilization of resources implied that interclan cooperation was not hindered by interclan rivalry. But because interclan cooperation advanced Genoa's economic prosperity (an *endogenous* change in a quasi-parameter), it intensified (given beliefs about clan motivations) the competition over political and economic dominance in the city. Fearing that any temporary decline in its relative power would constitute an opportunity that the other clan would take advantage of, clans became engaged in an "arms race" (leading to yet other *endogenous* changes in quasi-parameters): they bought land that they then fortified to dominate particular quarters, they established patronage networks, and they socialized their members to internalize loyalty to the clans 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 following 1154, attempts by the Emperor Frederick Barbarossa to regain de facto control over northern Italy weakened the link between political order and mutual deterrence. This particular external threat did not alter beliefs of the clan members of what other clans would do if the threat receded, but it did alter beliefs about the long-lastingness of the external threat, and this change not only lowered each clan's incentive to challenge the other militarily, but also made Genoa's political institutions self-enforcing in a wider range of situations than before. The result was that the Genoese clans mobilized their resources, acquired overseas commercial possessions, and expanded commercially. Through peace, Genoa's economic structure was transformed to one based on long-distance commerce.

Yet were the imperial threat to be weakened, this commercial expansion and structural transformation would undermine interclan mutual deterrence by making it self-enforcing for a smaller range of parameters. The higher level of economic prosperity (a change in a quasi-parameter), with concomitant gains to a clan for controlling the city, in the absence of an external threat, implied a smaller set of parameters for which mutual deterrence would have been self-enforcing.

In 1164 unexpected civil wars in Germany engaged the Emperor's attention. Genoa returned to its pre-1154 situation of low level of external threat. But the quasi-parameter of wealth was now higher than it was before, and with beliefs remaining stable, the previous mutual deterrence equilibrium between the clans was

[13] CDG 1936, vol. 1, no. 285. For the development of Genoa's political system consulate, see de Negri 1955 and Vitale 1955. Along with these consuls, other consuls responsible for the judicial system were also part of the administrative structure.

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 again when exogenous conditions changed. As noted by a twelfth-century Genoese chronicler, "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, 219–20). The fighting from 1189 to 1194 was particularly devastating and endangered the existence of the city.

These events in Genoa do not reflect only the influence of a shift in exogenous conditions. Rather, they reflect the fact that endogenous changes—increasing commercialization and prosperity, the clans' past investments in military ability and patronage, and, arguably, 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 prior to 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, now needing the assistance of Genoa's navy, had an interest in ceasing the civil war. By the promise of rewards and threat of war he 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 who was hired for a year to be Genoa's military leader, judge, and administrator and who 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 between them. The podestà's "threat" of assisting the other faction deterred each clan from attempting to control the city. Moreover, because the podestà was to receive a high wage at the end of his successful term, his threat was credible. For if one clan took control of the city, there was no reason for it to reward the podestà. Furthermore, this reward scheme made it in the podestà's interest not to alter fundamentally the balance of power between the factions. Hence he could credibly promise to be impartial and retaliate only against people who broke the law rather than against an entire clan. The selection of an incoming podestà was made by a committee of representatives from the city's various neighborhoods. It was large enough not to be dominated by any particular clan. The podesteria fostered interclan cooperation for a while and, thus, political stability and economic growth. It was a self-enforcing institution: The self-enforcing belief in the futility for any clan to gain political dominance using force deterred clans from trying. The belief that all clans could gain from cooperation without risking their economic position through unexpected military confrontation also motivated cooperation.

Yet, like the consular system, the podesteria was not reinforcing—indeed, it contained the seeds of its own destruction. Specifically, since the podesteria was based on balance of military strength between the clans and each clan wanted to be militarily prepared in case of need, it contained interclan rivalry but did not eliminate it. Each clan was still motivated to strengthen itself militarily vis-à-vis 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 in this period were both further manifestations of nonreinforcement 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 by assuming a common surname, usually that of the albergo's most powerful clan (Hughes 1978, 129–30). By the fifteenth century the city's politics and economics were dominated by about 30 alberghi, each containing from 5 to 15 lineages. With the alberghi, battle lines between clans hardened. Furthermore, the attempt of each clan to develop a patronage network and the access of all residents of the city to Genoa's overseas lucrative trade both implied that over time the nonnoble element of the city—the popolo—would acquire the resources, organization, and self-consciousness regarding their common interests to form a political faction that could disrupt the noble-controlled equilibrium.

Despite the peace of the podesteria, Genoa's institutions motivated clans to further invest resources in acquiring military ability aimed at attacking other clans, fortifying their residences, establishing patronage networks (thereby mobilizing the popoli), and indoctrinating their members to internalize the norms of revenge and adopting identities (through the alberghi) as clan members. In the short run, all these changes did not render the podesteria ineffective; it was still self-enforcing. But over time these changes caused Genoa's political structure to become self-enforcing for 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.

Venice. The history of Venice during its early days parallels that of Genoa. After an initial period of interclan cooperation, Venice's history was characterized by interclan rivalries aimed at capturing the office of the Doge (Lane 1973; Norwich [1977] 1989). Originally the Doge was a Byzantine official, but shortly after Venice was established in 679, the post became that of an elected monarch. For the next few hundred years, clans fought in Venice for control over the Doge's post. Similar to 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 from forming a political institution enabling cooperation. They responded to this opportunity by establishing a new

self-enforcing institution. At its center was the belief that each clan would join together to fight against a renegade clan that attempted to gain political dominance over the city and its economic resources. Arguably, this belief and the behavior it entailed fostered a common Venetian identity. In any case, the belief was sustained by a set of rules whose prescribed behavior was made self-enforcing by that belief. 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 fairly so all had a share in them regardless of clan affiliation. This allocative rule therefore did not give incentives for clans to increase their military strength to plan for interclan military conflict. Since the establishment of these rules began to occur when Byzantine and Muslim naval powers were on the decline and cooperation was most beneficial, Venetians were able to make the most of this opportunity.

Starting 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 it was established that a Doge should never act contrary to the advice of his councilors. The selection of the new Doge was entrusted to an official nominating committee to inhibit the ability to use a clan's political machine and popular support to influence the election. The nominating committee itself was selected and formed through an elaborate process that used both lots and delegations. The (partially random) process began in the Great Council, in which all adult nobles were eligible, so that all Venice's main clans had representatives. From this Council a committee of 30 was chosen by lot and the selection proceeded through an additional nine steps of delegation and selection by lot until the proposed candidate for Dogeship was brought before the Venetian assembly for approval. The importance of clans was reduced in this process by requiring that only one family member could be on any committee, and a person could not vote when a relative was being considered. The process itself was designed to reach a quick decision.

Similar, although less elaborate, systems were used for the selection of other officials. Their numbers were relatively large and their time in office relatively short, so that members of many clans could hold an office in a given period of time. Nomination 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 nomination committee. The conduct of all officials (including the Doge) was subject to scrutiny by committees to prevent unlawful gains.

That each clan had a stake in the implementation of these rules made self-enforceable the belief that each clan would join others to confront any clan that attempted to use military power to gain control over the city. But the rules and the associated beliefs were also reinforcing: They provided clans with few incentives to invest their resources in fortifying their residences or instilling in their members norms of loyalty to the clan rather than the city. By weakening the clans, Venice's republican magistracy over time increased the range of situations in which it was self-enforcing. This institution also prevented the endogenous formation of a political faction among nonnoble elements of the city, 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.[14]

Summary. During the eleventh and twelfth centuries, the self-enforcing institutions that emerged in Venice and Genoa were successful in fostering interclan cooperation. Yet they had distinct long-run implications.[15] In Genoa these institutions contained interclan rivalry but did not eliminate it. Each clan was motivated to militarily strengthen itself vis-à-vis the others, and most likely, individuals identified themselves more as clan members than as Genoese. Moreover, as the popoli grew in strength but were excluded from politics, they could disrupt coalitional governments among clans. Genoa's institutions eroded over time the range of situations in which they were self-enforcing, leading to their eventual collapse. In sharp contrast, Venice's institutions had reduced the political importance of clan structure. They discouraged the formation of clan alliances within the city and assured each clan that it would benefit from the political order and economic prosperity of Venice even if it did not belong to a clan alliance faction. Venice's institutions but not Genoa's were self-reinforcing.

Cleavage Structures in Nigeria and Estonia

The institution that concerns us in the paired comparison of Nigeria and Estonia is that of the *cleavage structure* within a polity.[16] In its classic formulation (Lipset and Rokkan 1967), it is stipulated that all societies divide themselves on a range of ascriptive, professional, class, and status criteria. Each of these divisions constitutes a social cleavage. A left–right dimension reflecting social class is a common cleavage for industrial societies. In many societies, however, people divide themselves politically based on tribal or nationality criteria. Cleavages are salient to the extent that individuals condition their behavior based on their category of membership on a cleavage dimension. In American cities, race is a salient cleavage since many voters condition their vote as members of a particular racial category; in Third Republic France, the religious/secular

[14] Several times this group had been extended to absorb emerging nonnoble families. The system therefore had the flexibility required for its perpetuation.

[15] The above discussion does not indicate the sources of these distinct institutions. Did the Genoese implement political organizations that in the long run were found to be devastating due to shortsightedness or some different initial conditions? Better theoretical understanding of organizational innovations and a more detailed historical analysis may provide the answer.

[16] This section is based on field research by Laitin in Nigeria and Estonia. For full developments of these arguments, see Laitin 1986, 1998.

cleavage was highly salient, again, as citizens conditioned their votes largely on whether they were regular Church goers. The relative salience of cleavages within a society and the number of cleavages that have any salience (along with the categories of membership on each dimension) constitute the cleavage structure.

Cleavage Structures as Institutions. Cleavage structure is usually thought of not as an institution, but more as a social reality. Yet it fits well within our definition of institutions. Consider first our criterion of an institution as being human-made. While schools of thought a generation ago held that people were given their social identities, it has become increasingly accepted that in fact people construct those identities. In the now-classic study by Thompson (1963), we know that the English "working class" had to be "made" through concerted social, cultural, and political activity. The creation of tribal cleavages in Africa or caste cleavages in India similarly has a strong human-made component (Chandra 2004; Laitin 1986). More important, the relative salience of cleavages—for example, that of religion in western Europe—is the result of social and political activity (Kalyvas 1996).

The elements that make up a cleavage structure include beliefs, organizations, and rules. Take rules. While categories such as "Jewish" and "African-American" and "worker" are constructed, there are usually clear rules concerning who can claim membership in these categories. Jews accept rules of maternal descent; in the United States an ounce of black blood is the rule for African-American membership; and salaried employment in blue-collar work is a rule-of-thumb for someone to qualify as a worker. As for organizations, categories of membership on cleavage structures often have preexisting organizations (e.g., churches, unions), but it is equally the case that organizations emerge to represent groups identified by salient cleavages (e.g., ethnic associations). And, as will shortly become clear, cleavage structures entail beliefs about the social world and how it is constituted.

As per our definition, the system of human-made elements that constitute the cleavage structure is exogenous to each individual whose behavior is influenced. People condition their political behavior in part on their self-identifications across social divisions or cleavages and on the perceived salience of those cleavages in political competition. Yet each individual cannot unilaterally legislate a new hierarchy of cleavages or a new set of categories of membership that define cleavages. Cleavage structures are therefore endogenous to societies but exogenous to any individual member of that society. And, finally, cleavage structures govern a transaction—political support or votes for group favors.

In studying cleavage structures and their implications for social peace and economic growth, many analysts take the ethnic/linguistic fractionalization (ELF) of a country's population to be parametric (Barro 1997; Easterly and Levine 1997). There is a clear plausibility to this—if there are two tribes in a society, one with 55% of the population and the other with 45%, we should expect a different form of conflict than if there were 100 tribes, each with 1% of the population. ELF is one algorithm designed to capture consequential differences in the ethnic arithmetic of a society. If ethnic demography is exogenous to political conflict, then ELF can be taken as one parameter governing the equilibrium supporting a particular societal cleavage structure.

There is a problem, however, in postulating ELF as a parameter in support of a particular cleavage structure. Suppose the equilibrium allows for slow processes of assimilation, or emigration, or pogroms. In these cases, ELF can change in value in the long term. We therefore postulate ELF as a quasi-parameter. We want to see whether changes in ELF, induced by the self-enforcing cleavage structure, reinforce or undermine that cleavage structure. In our paired comparison, we suggest that given parametric shift in ELF, the tribal-based cleavage structure in Nigeria is self-reinforcing but that the nationality-based cleavage structure in Estonia is self-undermining.

Application. The cleavage structure in Estonia and Nigeria shares two institutional elements: the organization of national (tribal) groups and everyday primordialist beliefs that supported this organization. Where institutional elements differ is that in Estonia the majority group has high social status that is recognized by the minorities, while in Nigeria the plurality group does not enjoy such status. The embeddedment of these cleavage structures in new states—states engaged in nation-building as well as management of their national economies—can be considered parametric. But the ELF index (that is, the categories of nation/tribe and the numbers in each category) is modeled here as quasi-parametric. Beliefs in the high status of Estonians give incentives, if voting results consistently favor Estonians, for non-Estonians to assimilate; over time these small changes in identity will affect the value of ELF. In Nigeria, there are no such incentives to assimilate, and therefore ELF is not likely to change in value. Our analysis foresees that in the context of political organization for elections (the transaction that is governed by the institution), a factor exogenous to the cleavage structure in Estonia (unitary political institutions) will transform Estonia's cleavage structure over time. Meanwhile, for elections in Nigeria, without any endogenous change in the quasi-parameter of ELF, shifts in exogenous political and economic institutions (moves from federalism to unitary government or reduction in the importance of oil) would not, because it is self-reinforcing, likely change its tribally based cleavage structure.

We now describe the cleavage structures. In newly independent Nigeria (1960), political parties were regionally based and catered to the needs and aspirations of the majority tribal group in their regions. In the Western Region, the Action Group catered to the interests of the dominant Yoruba population, and Yorubas largely associated themselves as supporters of that party. Within the Yoruba Region, factions within the Action Group represented the interests of subtribes of the Yoruba, associated with different ancestral cities. Similarly the NPC, the party of the Northern Region,

catered to Hausa interests; and the NCNC, the party of the Eastern Region, catered to Ibo interests. We can summarize the dominant cleavage structure of newly independent Nigeria as tribally based, with three principal groups dividing the political pie.[17]

In independent post-Soviet Estonia (1991), the dominant cleavage was based on nationality, one that pits a nationalist majority of Estonians against a largely disenfranchised but nationally conglomerated population of Russian-speakers (made up of Russians, Ukrainians, Belarusans, and Jews). The majority of Russian-speakers in Estonia are from families that migrated to Estonia after the Soviet occupation of 1940 and do not have command of the Estonian language. Most of these Russian-speakers did not qualify for citizenship in 1991 and therefore could vote only in local elections.

Nigeria and Estonia both have self-enforcing cleavage structures. Political leaders present platforms and candidate lists that reflect the interests of nationality constituencies and voters tend to respond to symbols and messages that speak to them as members of a particular tribal or nationality group. These cleavage structures are sustained by beliefs that have been dubbed "everyday primordialism" (Fearon and Laitin 2000). Primordialism is the belief that ethnic/nationality differences are biologically given and ultimately more important than any other possible identification when it comes to social, political, or economic transactions. Primordial beliefs of this sort are hardly universal. In the two cases here, they were created and sustained under previous political structures. British colonialism ruled "indirectly" through tribal chiefs, who were under salary by the British colonial state. These tribal chiefs were granted levels of authority they rarely achieved in the precolonial period, and Nigerians, to get a hearing from the British overrulers, had to petition through tribal authority structures. Thus colonialism played an important role in delineating tribal boundaries, clarifying tribal cleavages, and generating primordial beliefs.

Stalinist rule in the Soviet Union had a similar structure (Suny 1993). Each nationality was given its own political entity, and people's rights to position and property were dependent in large part on their nationality, as recorded on their passport. Soviet campaigns to change popular beliefs oftentimes only changed the public presentation of selves. Many Soviet citizens acted as if they were atheists but were secret believers; many Soviet citizens portrayed themselves as Stakhanovites who worked ceaselessly for the common good but privately stole from the state to sustain their families. But with the policy of *korenizatsiia* (cultivating local roots), the Stalinist system succeeded in creating a common knowledge of nationality. People not only accepted their nationality designations as real, but also believed that others believed their nationality to be real. An excellent example of this is in the Central Asian republics, where the Soviets named and codified nationality groups that were new and foreign to those who were assigned new nationalities (Allworth 1990). Yet two generations later, these designations became the bedrock of a set of newly independent national republics. Under British and Soviet overrule, tribal/nationality differences were codified, distributions were made in terms of tribal/national identities, and local political authorities who had monopoly representative rights for their tribal/nationality groups had an interest in cultivating beliefs that the cultures of their constituents were primordially given.

Nigeria. While tribal- and nationality-based cleavage structures have been self-enforcing in both Nigeria and Estonia, only in the Nigerian case can they be considered self-reinforcing. In Nigeria, the cleavage structure is self-reinforcing in that beliefs about the value of one's own tribal group, if anything, increases. Furthermore, institutions exogenous to the cleavage structure lend support to it. Federal institutions were built into the independence constitution. Political distributions were made based on formulae that returned federal funds to the original three regions. In 1967, the Eastern Region (whose leaders were opposed to the formula for the distribution of newly gained oil revenues) attempted secession, but lost a three-year bloody war fought against federal forces. Subsequent to the war, several minority tribes were given their own federal units (then called states). Each of the 12 states got a budget supported in large part on federally collected oil revenues. Since each state got a base allocation to cover the infrastructure of rule, smaller and smaller nationality groups grasped the incentives to demand their own states. By 1996, there were 36 separate states, almost all dominated by a single tribal group. Parallel to the demands for single tribe states, there have been in Nigeria increasing demands for the designation as a new local governing area for smaller (and sub-) tribal groups. For just a single example, in the city of Warri a communal war has pitted the Itsekiris who have controlled the local government against the Ijaws and Urhobos, both of whom want the city divided into separate local governing areas such that each of these groups can have control over a federally dispensed budget.

The federal system, in rewarding groups that can demonstrate a primordial claim to differentiate itself from the group that controls a particular state, supports the tribally based cleavage structure (Suberu 2001). Furthermore, with oil as the only part of the economy that provides substantial revenues, and with the federal authorities controlling those oil revenues, distributions from the center were of increasing relative value. Tribal groups that want more revenues need to claim primordial difference from the rulers of their state in order to be awarded a state of their own. With a state of their own, these groups have a claim on oil revenues for their own purposes. Through oil and federalism—exogenous to the cleavage structure—tribal identification gains increasing relevance, as these identifications

[17] Each of these groups has subgroups that seek community resources and would need to be described as part of the overall cleavage structure. Similarly, the three main tribes constitute only about half the Nigerian population, so a multitude of minority tribes is also an essential part of the cleavage structure. Later we discuss these elements within the three large tribal groupings and outside them.

are the key to statehood (and thus the value of ELF remains unchanged). More important, because tribal membership is key not only to state recognition but also to tribal group resources, tribe remains the dominant Nigerian cleavage. In this case federalism and oil are exogenous and parametric. But unless there are some endogenous processes undermining tribal membership rules (or promoting assimilation, doubtful given the institutional beliefs about status), the value of ELF as a quasi-parameter will not change, thus making the Nigerian cleavage structure self-reinforcing. This is what we mean by an institution surviving under a wider range of situations. It is the stronger boundaries between tribal groups, reinforced by beliefs about status, that stabilize ELF as a quasi-parameter, thereby making the primordially based cleavage structure in Nigeria self-reinforcing.

Estonia. Although *korenizatsiia* was abandoned as official policy by the mid-1930s, Estonia (coercively brought into the Union in 1940 based on von Ribbentrop's secret pact with Molotov) under Stalin's Bureau of Nationalities inherited many features of that policy. Estonians were in an institutional sense the "owners" (called "titulars") of the Estonian Soviet Socialist Republic. All party and state documents, education, and the media in Estonia would be transmitted either in Russian (the language of Russia, the "elder brother" of all titular peoples) or in Estonian. Korenizatsiia institutions sustained the belief throughout the population of everyday primordialism that Estonians were forever and primarily Estonians, even if the Soviet Union was a suprarepublic that made possible the "brotherhood of the peoples." In the standard country study written by an Estonian, the notion that a Russian could be an Estonian was not even considered (Raun 1987). Nationality boundaries are popularly conceived of as consequential for political loyalties and not subject to change.

With the breakup of the USSR, with titulars in the majority, Estonian parliamentarians declared Estonia to be a unitary state and wrote citizenship laws that effectively disenfranchised (for national elections) the 30% of the population that was Russian-speaking with roots outside of Estonia. At first, these new institutions united the Russian-speaking population into an opposition front, with some leaders articulating a goal of reunification with the Russian Federation. In this sense, the primordial cleavage dividing Estonians from Russian-speakers was self-enforcing.

But the cleavage structure in Estonia is not self-reinforcing. The relatively high status of Estonian language and culture in the eyes of the Russian-speakers is a belief that encourages Russian assimilation into Estonian culture, potentially changing the value of ELF. Exogenous parameters such as the opportunities as citizens of Estonia to take part in European institutions, the unitary political institutions, and the chances to participate in those institutions with the learning of the Estonian language all gave incentives, especially to young Russian-speakers, to assimilate into Estonian society. In the short term, the quasi-parameter of the national demography or ELF of the population is only marginally changing. Estonian is becoming (for some cross-national interactions) the language of internationality communication between Estonians and Russians. At first, only Russian-speakers with a low cost for passing (e.g., those who learn languages easily) become Estonians. In the longer term, there could well be a new generation of Russian-speakers fluent in Estonian. In consequence, the cultural differences between Estonians and (descendants of) Russian-speakers will begin to narrow. Russians can become Estonians, or Russian-Estonians, thereby changing the value of ELF in the longer term. As the country's population through assimilation becomes decreasingly divided by language and nationality, the Russian/Estonian cleavage will decline in the realms for which such identifications are today significant. We can thus project institutional undermining of the dominant cleavage in the longer term, perhaps giving way to a class-based dominant cleavage. In this case, then, the ethnic cleavage structure in Estonia is self-enforcing but not self-reinforcing.

FORMAL REPRESENTATION OF INSTITUTIONAL REINFORCEMENT

Repeated game theory postulates that the same stage game is repeated each period. As such, it seems less promising for the study of institutional dynamics than dynamic game theory. Our judgment, however, is that repeated game theory seems to capture the way that people view their environment and make decisions. The information requirements and computational complexities of dynamic games are unrealistically demanding as a basis for a general theory of institutional change. Accordingly, we model endogenous institutional dynamics using the lens provided by repeated game theory.

The remainder of this section contains a formal representation of a game in which there is the possibility of endogenous shift in one of the parameters of the game (the payoffs).[18] Our purpose here is not to give an account of institutions but, rather, to illustrate how quasi-parameters and reinforcement processes can be incorporated into standard repeated game-theoretic models. Yet, to illustrate the generality of the discussion, we relate it to our empirical analyses.

We represent the institutions generating behavior in a particular transaction as a game and shared beliefs regarding self-enforcing behavior in it. The specification of a game and beliefs in it thus constitutes a statement of our thesis regarding the relevant institution. Indeed, it is a statement of our conjecture about the perception of the players regarding the situation (Rubinstein 1991) and the beliefs they hold about each others' predicted behavior. This game-theoretic

[18] The force of our argument regarding the importance of self-enforcing and undermining processes is not limited to the particular game structure or equilibrium refinement. Rather, our argument rests on the difficulties that individuals normally face while having to think their way through strategic situations.

framework thereby makes explicit the parameters delineating the extent of self-enforceability of various beliefs. Building on this framework, we can study institutional dynamics by combining what the analyst understands about the situation—particularly regarding processes that reinforce or undermine (quasi-) parameters—with a conjecture regarding what the decision makers understand, know, and observe.

To illustrate the implications of this formulation, consider the infinitely repeated prisoners' dilemma (PD) game presented in the Appendix. 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 *c*,*c* in equilibrium over repeated play).[19] There are four parameters in this game: the cooperative payoff for each player (b_t), the sucker's payoff (k), the additional payoff for defecting while the other player is cooperating (e), and the discount factor (δ). In this representation, however, b_t is, for illustrative purposes, a quasi-parameter.

We change one assumption in the standard repeat-play PD model where mutual cooperation has neutral feedback. Instead, we allow for three possibilities—neutral, positive, and negative feedback—which correspondingly lead to neutral, positive, and negative self-reinforcement (undermining). In a positive feedback situation, the payoff b after any *c*,*c* outcome increases by ϵ for the next round of play, thereby 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, thereby undermining the institution. The cooperative payoff changes conditioned on the outcome in the previous playing of the game. In the first case of positive reinforcement, over time the range of δ for which *c*,*c* will be self-enforcing increases. The institution of cooperation is not only self-enforcing but self-reinforcing: It is an equilibrium in the short run that, in the long term, would be so for a wider range of discount factors or other parameters. Conversely, in the case of undermining, cooperation is only self-enforcing but not self-reinforcing, as, over time, the range of δ for which *c*,*c* will be self-enforcing decreases. 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 our formalization, reinforcement and undermining processes are not dependent on actors' knowledge of the feedback mechanism. Yet, who possesses this knowledge determines these processes' institutional ramifications. Consider first (Case 1) a situation in which the actors are fully aware of the reinforcing (undermining process). In this case, positive reinforcement 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 would never be an equilibrium outcome due to unraveling.

In reality, however, other responses to foreseen undermining process are often possible and the study of Genoa reflects two of them. In Genoa, cooperation led to undermining by increasing wealth and hence the temptation to capture it. Each Genoese clan was therefore motivated to cooperate with other clans only to the extent to which its gains from the 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. The second type of response that Genoa's history illustrates is organizational and constitutes an institutional refinement. In 1194 the mutual deterrence equilibrium was no longer self-enforcing but its costs to both clans increased due to the Emperor's intervention threats. The response was organizational: the introduction of the podestà, an organization designed to restore mutual deterrence and cooperation and that reflects a process of learning.

Reinforcing and undermining processes are not necessarily recognized by the relevant actors (Case 2). In the PD game, ignorance of undermining would imply cooperation for several periods until the actors 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 actors will not effectively respond to it. Tribes in Nigeria may have been intentionally given the right to establish states to weaken the dominant rebellious group in the Eastern Region. But the subsequent proliferation of states and the resulting reinforcement of the cleavage system were probably not foreseen. Nonetheless, because the cleavage system is self-reinforcing, there is no endogenous process to reverse the trend. Here the observed implications of an unforeseen and probably undesired reinforcing process has not led to institutional refinement and redesign of the institution to reach an objectively desirable outcome.

It is also often the case that those who observe a process of undermining have little incentive to reveal it to others. Such one-sided knowledge regarding undermining would lead to collapse of the previous institution once the one who possesses that knowledge begins acting in a manner revealing it. It then can be followed by institutional refinement and redesign aimed

[19] In asserting that the players are engaged in the prisoners' dilemma game we already assert 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 that captures the prisoners' beliefs that cooperation lessens punishment. Potentially relevant organizations like the Mafia are assumed not to exist. Hence, the game assumes away the possibility of beliefs that a prisoner who defected would be penalized by such an organization. Similarly, the analysis assumes away the possible influence of norms such as that of honor among thieves that the prisoners may have internalized prior to their arrest. Such norms, if they had been internalized by the prisoners, would have affected their willingness to cooperate or defect.

at restoring a desired outcome given new knowledge regarding the situation.[20]

Our purpose in this section was not to provide a full game-theoretic analysis of institutions and their dynamics. Here we just seek to illustrate formally how the notion of quasi-parameters can be represented and to suggest ways that dynamic game theory might develop within the framework of infinitely repeated games and the subgame-perfection solution concept.

CONCLUDING REMARKS

This paper has examined—through analysis of reinforcing processes—why and how the behavior induced by self-enforcing institutions influences their long-term survival. 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. And institutional behaviors influence rates of institutional change directly, 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 and then examining the implied reinforcing processes. One can conclude by examining the long-term implications of these processes for that institution's endogenous rate of change.

Our approach is game-theoretic in orientation. However, we have shown how the techniques of rational choice institutionalism can be used to address the concerns of historical institutionalism in previously unappreciated ways: Properly understood, the two are *not* mutually contradictory, contrary to the portrayals of the discipline as one of a war of paradigms. Take, for example, "Tocqueville's revenge" (Levy 1999) now as a case of a self-undermining process. By acceding to state protection in earlier periods, small business associations (an institutional element of the French political economy) atrophied. This change decreased the range of situations in which development based on cooperation between the government and a "collective business partner" could be in the equilibrium set for policymakers. Where Pierson might portray this study of Tocqueville's revenge an example of increasing returns for state dirigisme, we see the marginal shift in a single quasi-parameter—number of local business associations—that slowly undermined an institutional equilibrium. Put this way, our conceptual apparatus allows institutionalists better to isolate the factors—as Levy did in a more discursive way—fostering institutional change.

The example of the one-party systems in Sweden and Italy provides an illustration of two institutions that were differentially able to recover from an external shock. The lesson here is that apparently similar institutions (dominant parties) can have very different elements and are therefore differentially resistant to exogenous shock. Meanwhile Ertman's analysis of one critical juncture—the timing of entry into the world system—is clearly an exogenous parameter that does not shift in value as the country develops. By isolating institutional elements, differentiating parameters from quasi-parameters, and differentiating exogenous and endogenous processes of change, our game-theoretic framework allows for better specification of historical institutionalist findings.

But our analysis, taking historical institutionalist concerns to focus on the processes of change, can enrich state-of-the-art rational choice institutionalism as well. For example, Bueno de Mesquita et al. (2003) analyze political institutions based on how long their leaders survive in office. Their key parameter is W/S (the ratio of the size of the winning coalition to the size of the entire "selectorate," i.e., those who have a role in choosing a new leader). The lower the ratio, the longer members of the winning coalition will remain loyal to their leader. However, an implication of loyalty and the concomitant longevity of rule is that members of S (who are not in W) will have an incentive to emigrate (as analyzed in chap. 8). Over time, this emigration will raise the value of W/S, which can undermine loyalty and make the political institutions more sensitive to external shock. In our terms, W/S is a quasi-parameter, and as S lowers in value as an implication of the equilibrium of long survival of rulers, dictatorships become self-undermining.[21]

There are several extensions called for with our approach to the endogenous shift in institutions, both methodological and substantive. Methodologically, we relied 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 our formal model. Our second methodological challenge is to subject contextually based game-theoretic analysis of institutions to statistical test. Unless the observable implications of our models are statistically examined over a range of cases that were not from the set of cases from which we developed our theory, 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 can serve as a test of the model's validity. For example, our model of the two Italian

[20] For historical examples of relevant actors adopting new and/or refined institutions because they learned or foresaw the undermining processes at work in their self-enforcing institutions, see Greif et al. 1994.

[21] For a similar dynamic model analyzing historical change, see Bueno de Mesquita 2000.

cities has the observable implication that over time in Venice, there would be more interclan exogamy than in Genoa. If this proves to be correct, it would help overcome charges of tautology. Furthermore, statistical tests will allow us to assess the relative importance of endogenous vs. exogenous sources of institutional change. Third, our analysis emphasized the importance of quasi-parameters but did not explore the features of institutions that foster reinforcing or undermining changes in quasi-parameters in various situations.

Substantively, the particular approach to the study of self-enforcing institutions presented above can be extended to examine why and how self-enforcing institutions influence the *direction* of institutional change. Such institutions influence the direction of institutional change through the persistence of their institutional elements—their rules, organizations, beliefs, and norms. When an institution ceases to be self-enforcing and no longer generates a particular pattern of behavior, its institutional elements provide some of the initial conditions in the process leading to a new institution (Greif n.d., Chapter 7). The direction of institutional change is thus path-dependent. This phenomenon was recognized by Swidler (1986) in her notion of culture as a "tool kit" enabling reconstitution of society in troubled times. Greif (1994) has shown how the institutional elements of medieval trader/agent transactions in Genoa were incorporated into its modern trade institutions. And in this paper (following Greif 1998), we have seen how the clans, beliefs, and norms that prevailed in Genoa prior to the introduction of the podesteria were incorporated into this new institution.

Our approach also invites the testing of conjectures that relate to our model. For example, our vignette of Genoa suggests that initially institutions tend to reinforce themselves (assuming they are self-enforcing) but as time passes undermining processes assert themselves, suggesting a stochastic life cycle of institutions.[22] This initial reinforcement reflects the role of institutions in encompassing information and their influence on belief formation. Beliefs shared by members of the society about how other individuals would behave in various contingencies are an institutional element. Each individual, however, faces some uncertainty as to whether such behavior will or will not be followed and to what effect. Basing one's actions on the beliefs about what others will do is not foolproof. Others' actions are not *ex ante* known with certainty, and as stressed above, the many environmental factors influencing others' behaviors are not directly observable. The *ex ante* expected value of goal-oriented behavior may be high, but *ex post* these strategies could still fail. When, *ex post*, these behaviors work, the uncertainty is resolved—this is what is meant by self-enforcement through the confirmation of beliefs by actors as to what other actors are likely to do—and the value of continuing to use them is higher than it was *ex ante*. The very 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. Hence, it is more likely to be followed. Over the longer term, however, marginal changes in quasi-parameters have their impact. They can be self-reinforcing, as with city identities in Venice supporting the Doge and ELF in Nigeria making tribal-based cleavages even more salient than in earlier periods. But these marginal changes can equally be self-undermining.

However such extensions and conjectures work out, we have provided a framework, with the introduction and elaboration of the concepts of quasi-parameters and institutional reinforcement, to integrate the study of self-enforcing institutions with that of endogenously induced institutional change.

[22] This is consistent with an observation in Pierson 2000 (253).

APPENDIX: FORMAL MODEL OF INSTITUTIONAL REINFORCEMENT

Take an infinitely repeated PD game in which the period $t = 0, 1, \ldots$, stage-game payoffs are

$$\begin{array}{cccc} & c & & d \\ c & b_t, b_t & & -k, b_t + e \\ d & b_t + e, -k & & 0,0 \end{array} \qquad \textbf{(1)}$$

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

Definition. Cooperation has a positive (negative, neutral) reinforcement if play of (c,c) in period t implies that $b_{t+1} - b_t > (<, =)\ 0$. Standard repeated PD models take cooperation to have neutral reinforcement. To simplify the analysis, we assume that the change in cooperation payoffs under any reinforcement mechanism is fixed over time.

Assumption. For all t, $b_{t+1} - b_t = \varepsilon$ with $\varepsilon > (<, =)\ 0$ under positive (negative, neutral) reinforcement. In what follows, our equilibrium notion is SPNE. Since the analysis is not particularly complex, we are somewhat informal to avoid complicating notation and terminology.

Case 1. Knowledge about Reinforcement

Consider the case where 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

$$\delta \geq \frac{e}{b_\tau + e}. \qquad \textbf{(2)}$$

On the other hand, suppose there is positive reinforcement. Recall that $\varepsilon \equiv b_{\tau+1} - b_\tau > 0$ under Nash reversion. If players follow Nash reversion, then on the equilibrium path their payoffs are strictly *larger* than $b_\tau + (b_\tau + \varepsilon)[\delta/(1-\delta)]$. On the other hand, deviating gives $b_\tau + e$. Hence cooperation is

incentive compatible if $e \leq (b_\tau + \varepsilon)[\delta/(1-\delta)]$, which rewrites as

$$\delta \geq \frac{e}{b_\tau + e + \varepsilon}. \quad \textbf{(3)}$$

Since $\varepsilon > 0$, the RHS of (2) is strictly smaller than the RHS of (2), which proves the claim. ■

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

Proof. Straightforward by backward induction given that payoffs from mutual cooperation decrease by ε every period if players have cooperated in previous periods. ■

So the institution of cooperation can only be self-enforcing under neutral or positive reinforcement. Moreover, under positive reinforcement, the institution is positively reinforcing since the RHS of equation (2) is decreasing over time (due to b_t increasing) and hence the equilibrium holds for a larger range of δ over time. On the other hand, by similar reasoning, the institution is neither positively nor negatively reinforcing under neutral reinforcement—the range of δ for which it is self-enforcing is exactly the same in any period t.

Case 2. Ignorance about Reinforcement

Now consider the case where players are not aware of the reinforcement mechanism. So 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 $b_\tau + e \leq [b_\tau/(1-\delta)]$ or, equivalently, if and only if

$$\delta \geq \frac{e}{b_\tau + e}. \quad \textbf{(4)}$$

The RHS of (4) is strictly decreasing in b_τ. Hence, if cooperation has positive reinforcement, then 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. On the other hand, if cooperation has negative reinforcement, then 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.

REFERENCES

Allworth, E. A. 1990. *The Modern Uzbeks*. Stanford, CA: Hoover Institution Press.

Annali. 1190. *Annali Genovesi di Caffaro e dei suoi Continuatori, 1099–1240, 1923–1929*. 4 vols. Trans. Ceccardo Roccatagliata Ceccardi and Giovanni Monleone. Genoa: Municipio di Genova.

Aoki, Masahiko. 2001. *Comparative Institutional Analysis*. Cambridge, MA: MIT Press.

Aumann, Robert J., and Adam Brandenburger. 1995. "Epistemic Conditions for Nash Equilibrium." *Econometrica* 65 (5): 1161–80.

Barro, Robert. 1997. *Determinants of Economic Growth*. Cambridge, MA: MIT Press.

Bueno de Mesquita, Bruce. 2000. "Popes, Kings, and Endogenous Institutions: The Concordat of Worms and the Origins of Sovereignty." *International Studies Review* 2 (2): 93–118.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson, and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge, MA: MIT Press.

Calvert, Randall L. 1995. "Rational Actors, Equilibrium, and Social Institutions." In *Explaining Social Institutions*, ed. Jack Knight and Itai Sened. Ann Arbor: University of Michigan Press, 57–93.

CDG. 1936, 1938. *Codice Diplomatico della Repubblica di Genova dal MCLXIIII [sic] al MCLXXXX [sic]*. Vols. I, II, ed. Cesare Imperiale di Sant'Angelo. Rome: Fonti per la Storia d'Italia.

Chandra, Kanchan. 2004. *Why Ethnic Parties Succeed*. Cambridge: Cambridge University Press.

Collier, David, and Ruth Collier. 1991. *Shaping the Political Arena*. Princeton, NJ: Princeton University Press.

David, Paul. 1994. "Why Are Institutions the 'Carriers of History'?: Path-Dependence and the Evolution of Conventions, Organizations and Institutions." *Structural Change and Economic Dynamics* 5 (2): 205–20.

de Negri, Teoflio Ossian. 1986. *Storia di Genova*. Florence: G. Martello.

DiMaggio, Paul J., and Walter W. Powell. 1991. "Introduction." In *The New Institutionalism in Organizational Analysis*, ed. Paul J. DiMaggio and Walter W. Powell. Chicago: University of Chicago Press.

Durkheim, Emile. [1895] 1950. *The Rules of Sociological Method*. New York: Free Press.

Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy." *Quarterly Journal of Economics* 112 (November): 1203–50.

Epstein, Stephan. 1996. *Genoa & the Genoese, 958–1528*. Chapel Hill: University of North Carolina Press.

Ertman, Thomas. 1997. *Birth of the Leviathan.* New York: Cambridge University Press.

Fearon, James D., and David D. Laitin. 2000. "Violence and the Social Construction of Ethnic Identities." *International Organization* 54 (Autumn): 845–77.

Fudenberg, D., and D. Levin. 1998. *The Theory of Learning in Games*. Cambridge, MA: MIT Press.

González de Lara, Yadira. 2004. "The State as an Enforcer in Early Venetian Trade: A Historical Institutional Analysis." Memo. Department of Economic Analysis, University of Alicante.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83 (3): 525–48.

Greif, Avner. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102 (5): 912–50.

Greif, Avner. 1995. "Political Organizations, Social Structure, and Institutional Success: Reflections From Genoa and Venice during the Commercial Revolution." *Journal of Institutional and Theoretical Economics* 151 (4): 734–40.

Greif, Avner. 1997. "Microtheory and Recent Developments in the Study of Economic Institutions through Economic History." In *Advances in Economic Theory*, ed. David M. Kreps and Kenneth F. Wallis. Cambridge: Cambridge University Press, Vol. II.

Greif, Avner. 1998. "Self-Enforcing Political Systems and Economic Growth: Late Medieval Genoa." In *Analytic Narratives*, ed. Robert H. Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. Princeton, NJ: Princeton University Press.

Greif, Avner. n.d. *Institutions: Theory and History.* Cambridge: Cambridge University Press. Forthcoming.

Greif, Avner, Paul R Milgrom, and Barry R. Weingast. 1994. "Coordination, Commitment and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102 (4). (August) 745–76.

Hall, Peter A., and Rosemary C. R. Taylor. 1996. "Political Science and the Three New Institutionalisms." *Political Studies* 44 (5): 936–57.

Hall, Peter A., and Rosemary C. R. Taylor. 1998. "The Potential of Historical Institutionalism: A Response to Hay and Wincott." *Political Studies* 46 (5): 958–62.

Herlihy, David. 1969. "Family Solidarity in Medieval Italian History." In *Economy Society, and Government in Medieval Italy, Essays in Memory of Robert L. Reynolds*, ed. David Herlihy, R. S. Lopez, and V. Slessarev. Kent, OH: Kent State University Press, 173–84.

Hughes, Diane. 1978. "Urban Growth and Family Structure in Medieval Genoa." In *Towns in Societies*, ed. Philip Abrams and E. A. Wrigley. Cambridge: Cambridge University Press.

Ikenberry, John. 1994. "History's Heavy Hand." Presented at conference on The New Institutionalism, University of Maryland.

Kalai, E., and E. Lehrer. 1993. "Rational Learning Leads to Nash Equilibrium." *Econometrica* 61 (September): 1019–45.
Kalai, E., and E. Lehrer. 1995. "Subjective Games and Equilibria." *Games and Economic Behavior* 8: 123–63.
Kalyvas, Stathis N. 1996. *The Rise of Christian Democracy in Europe*. Ithaca, NY: Cornell University Press.
Katznelson, Ira. 1997. "Structure and Configuration in Comparative Politics." In *Comparative Politics: Rationality, Culture and Structure*, ed. M. Lichbach and A. Zuckerman. New York: Cambridge University Press.
Krasner, Steven. 1984. "Approaches to the State." *Comparative Politics* 16 (January): 223–46.
Laitin, David D. 1986. *Hegemony and Culture*. Chicago: University of Chicago Press.
Laitin, David D. 1998. *Identity in Formation*. Ithaca, NY: Cornell University Press.
Lane, Frederic C. 1973. *Venice: A Maritime Republic*. Baltimore: Johns Hopkins University Press.
Levy, J. 1999. *Tocqueville's Revenge*. Cambridge, MA: Harvard University Press.
Lipset, S. M., and Stein Rokkan. 1967. "Cleavage Structures, Party Systems, and Voter Alignments: An Introduction." In *Party Systems and Voter Alignments*, ed. S. M. Lipset and Stein Rokkan. New York: Free Press.
Lopez, Robert S. 1976. *The Commercial Revolution of the Middle Ages, 950–1350*. London: Cambridge University Press.
Marimon, Ramon. 1997. Learning from Learning in Economics. In *Advances in Economic Theory*, Vol. I, ed. David M. Kreps and Kenneth F. Wallis. Cambridge: Cambridge University Press, 278–315.
Muthoo, Abhinay, and Kenneth A. Shepsle. 2003. "Agenda-Setting Power in Organizations with Overlapping Generations of Players." Typescript. Harvard University.
Moriguchi, Chiaki. 1998. "Evolution of Employment Systems in the US and Japan: 1900–60. A Comparative Historical Analysis." Manuscript. Stanford University.
North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
Norwich, John Julius. [1977] 1989. *A History of Venice*. New York: Vintage Books.
Pierson, Paul. 1994. *Dismantling the Welfare State?* Cambridge: Cambridge University Press.
Pierson, Paul. 2000. "Increasing Returns, Path Dependency, and the Study of Politics." *American Political Science Review* 94 (June): 251–67.
Pierson, Paul, and Theda Skocpol. 2002. "Historical Institutionalism in Contemporary Political Science." In *Political Science: State of the Discipline*, ed. Ira Katznelson and Helen V. Milner. New York: W. W. Norton, 693–721.
Raun, Toivo. 1987. *Estonia and the Estonians*. Stanford, CA: Hoover Institution Press.
Rubinstein, Ariel. 1991. "Comments on the Interpretation of Game Theory." *Econometrica* 59 (July): 909–24.
Rubinstein, Ariel. 1998. *Modeling Bounded Rationality*. Cambridge, MA: MIT Press.
Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Schotter, Andrew. 1981. *The Economic Theory of Social Institutions*. Cambridge: Cambridge University Press.
Simon, Herbert. 1976. *Administrative Behavior*. 3rd ed. New York: Macmillan.
Suberu, Rotimi T. 2001. *Federalism and Ethnic Conflict in Nigeria*. Washington, DC: USIP Press.
Suny, Ronald G. 1993. *Revenge of the Past*. Stanford, CA: Stanford University Press.
Swidler, Ann. 1986. "Culture in Action." *American Sociological Review* 51 (April): 273–86.
Tabacco, Giovanni. 1989. *The Struggle for Power in Medieval Italy*. Cambridge: Cambridge University Press.
Thelen, Kathleen. 1999. "Historical Institutionalism in Comparative Politics." *Annual Review of Political Science* 2: 369–404.
Thompson, E. P. 1963. *The Making of the English Working Class*. London: Victor Gollancz.
Ullmann-Margalit, Edna. 1977. *The Emergence of Norms*. Oxford: Clarendon Press.
Vitale, V. 1955. *Breviario della Storia di Genova*, 2 vols. Genoa: Societa Ligure di Storia Patria.
Weingast, Barry R. 1996. "Political Institutions: Rational Choice Perspectives." In *A New Handbook of Political Science*, ed. Robert Goodin and Hans-Dieter Klingemann. Oxford: Oxford University Press.
Young, H. Peyton. 1998. *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press.