Comparing Theories of Institutional Change

Christopher Kingston[1] and Gonzalo Caballero[2]

August 11, 2006

Abstract

This paper compares some recent theoretical approaches to conceptualizing institutional change, in an effort to clarify areas of consensus and disagreement regarding the causes, process, and outcomes of institutional change. We do not attempt to build a "new" theory, and our survey is not intended to be comprehensive. Among the theories we discuss, some emphasize spontaneous evolutionary processes, while others emphasize deliberate design, or combinations of evolution and design. We differentiate a variety of approaches to conceptualizing the interaction between formal and informal rules. We discuss recent theories based on the "equilibrium view" of institutions, and theories emphasizing the role of cognition and bounded rationality. We also consider theoretical explanations for institutional inertia and path-dependence.

[1] Corresponding author. Economics Department, Amherst College, Amherst MA 01002. cgkingston@amherst.edu
[2] Faculty of Economics, University of Vigo, Lagoas Marcosende, 36310 VIGO, Spain. gcaballero@uvigo.es

1. Introduction

Recent work in transition economics, economic history, and economic development has highlighted both the importance and complexity of institutional change, and has led to some significant advances in our understanding. The relevant literature, however, is vast and diffuse, and plagued by a profusion of terminology, much of which is used in different ways by different authors. There is no consensus on how to conceptualize either institutions themselves or the process of institutional change.

This paper compares some recent theories of institutional change. We do not attempt to build a new theory or to set out how people "should" conceptualize institutions or institutional change. Rather, our goal is to try to map out some relationships among existing theories. Also, the paper is not intended as a comprehensive survey. In particular, we do not discuss historical institutionalism in sociology and political science (Thelen 2004; Campbell 2004), although many related issues arise in those literatures. The key issues we are interested in have to do with:

- The *causes* of institutional change: What are the effects of exogenous shocks and endogenous processes in bringing about change, and which is more important, in the short run and the long run?

- The *process* of institutional change: under what circumstances is change deliberate or spontaneous, sudden or gradual, a cooperative venture or an outcome of conflict? What are the sources of inertia which make institutions relatively durable over time? What is the role of politics? What is the role of "bounded rationality"? How should we think about the interaction between formal and informal rules?

- The *outcome* of institutional change: Under what circumstances will efficient institutions tend to emerge? When there are multiple equilibria, how are particular equilibria selected? When, and how much, does history matter?

The paper proceeds by outlining five broad approaches to conceptualizing institutional change: Evolutionary theories (section 2.1); theories which envisage a hierarchy of rules (section 2.2); theories which emphasize the interaction between formal and informal rules (section 2.3); the "equilibrium view" of institutions (section 2.4); and cognition-based theories (section 2.5). Our goal is clarify how these theories agree or differ with respect to the three sets of issues discussed above: the causes, process, and outcomes of institutional change. Section 3 discusses why institutions often exhibit considerable resistance to change. Section 4 concludes.

2. Theories of institutional change

Before we can discuss institutional change, we must define what we mean by "institutions". Unfortunately, different authors use different definitions of institutions, and this naturally influences their views of institutional change. The most commonly accepted definition is that institutions are the rules of the game in a society, together with their enforcement arrangements (North 1990). They include both formal rules such as laws and constitutions, and informal rules such as conventions and norms,[3] and are "humanly-devised", in the sense that they are a product of social interactions among people (thus, technological constraints like the "laws" of physics are not institutions).

For the time being, therefore, let us adopt this definition (we will discuss the implications of an alternative definition in section 2.4). If institutions are sets of rules, the question of institutional change becomes: how do the rules change?

[3] North excludes "organizations" – states, firms, tribes, families – from his definition of institutions. Implicitly, he treats organizations as entities within which collective action and agency problems have been solved, so that they can be treated as single actors: for North, organizations are the "players" of the game. Many other authors, however, include organizations within their definition of institutions: in the simplest sense, organizations are simply groups of individuals who interact with each other repeatedly in pursuit of common goals. The internal interactions within organizations may be governed by formal rules (such as a faculty handbook), or informal rules (no eating dinner in front of the TV). Certainly, studying the internal rules within organizations such as firms, communities and states is a central concern of much of institutional analysis.

2.1 Decentralized institutional change (Williamson 2000; Sugden 1989; Knight 1995)

We first consider a strand of research in which institutional change is a relatively minor issue. Transactions cost economics (TCE) argues that in many interactions, "transaction costs" arise because of the bounded rationality and opportunism of the transacting parties. Depending on the attributes of the particular transaction of interest, particular sets of rules ("governance structures") will be able to govern this transaction more efficiently than others. TCE *assumes* that these institutional forms (those which "minimize transactions costs") will tend to be observed - that is, that institutions will emerge so as to achieve an optimal "match" with a particular transaction: what Williamson refers to as the "discriminating alignment" hypothesis.

The basic cause of institutional change is a change in the exogenous parameters which affect transaction costs, such as production or monitoring technology. If existing institutions are no longer efficient following such a change, then new, more efficient institutional forms will gradually emerge.

Implicitly, the process of institutional change envisaged is one in which competitive pressure weeds out inefficient forms of organization, as originally suggested by Alchian (1950), because those who choose efficient institutions will realize positive profits, and will therefore survive and be imitated. However, the process of institutional change is of relatively minor interest, since the outcome of the change is determinate: the most efficient institutions will win.

Although bounded rationality is a key explanation for the existence of institutions in TCE, it does not play a major role in institutional change. Even if players introduce new institutions entirely at random, more efficient institutions will drive out less efficient ones through the evolutionary-competitive process (if people do try to design and imitate efficient institutions, this would simply accelerate the process). This competitive pressure ensures that bounded rationality will not impede the selection of efficient institutions.

This approach is "an empirical success story" (Williamson 2000: 607) which has proved fruitful in explaining many observed characteristics of exchange relationships. The scope is necessarily

limited, however, to situations in which competition among institutional forms can plausibly operate to weed out inefficient rules. For example, it can more readily be applied to a choice of contractual forms in an industry with many firms than to the choice of a constitution for a state.

In addition, although Alchian explicitly recognized the potential for this process to arrive at a local rather than a global optimum, much of the subsequent literature has ignored this possibility. Yet multiple equilibria can arise if, for example, the payoff to choosing a particular institutional form depends on the frequency of its adoption in the population overall (for example, credit cards are an efficient means of payment, because they are widely used, but they are widely used because they are an efficient – and therefore cheap and convenient – means of payment). The TCE approach therefore cannot explain why countries with similar technologies may use different institutions to govern apparently similar transactions; why inefficient institutions often seem to persist; or why less successful societies often fail to adopt the institutional structure of more successful ones.

Other evolutionary theories do consider the potential for multiple possible sets of self-enforcing rules (multiple "conventions"). The question of institutional change then revolves around how particular conventions are selected to begin with.

One possibility is that equilibrium selection might be the arbitrary result of historical accidents. However, there are several reasons why some rules may be more likely to emerge than others. Sugden (1989) argues that in novel situations, people wishing to coordinate their strategies will tend to coordinate on rules which are analogous to rules with which they are already familiar, such as the rule "first come, first served" rule for assigning property rights, variants of which can be applied to many situations.

Knight (1995) argues that different sets of rules often have different distributional consequences, so different actors may favor the emergence of different rules. In the period before rules relating to some interaction have been firmly established, individual actors engage in decentralized bargaining over which rule to adopt in their individual interactions. If some kinds of actors have

greater bargaining power than others,[4] this may affect the kind of rule that ultimately becomes widely-used by the society overall.

When there are multiple equilibria, a central concern is that it may be possible for inefficient equilibria to arise and persist. Unlike the TCE approach, neither Sugden nor Knight argue that efficient rules are necessarily likely to emerge. Instead, institutional change exhibits "path-dependence", in the sense that initial conditions and historical accidents can have a lasting impact on the institutions which emerge.

2.2 Hierarchies of Rules (Libecap 1989, Ostrom 2005)

Many authors treat institutional change not as a decentralized and spontaneous process, but rather as a centralized, political process in which the state specifies formal rules (or "property rights") and individuals and organizations engage in conflict and bargaining to try to change these rules for their own benefit.

One example of this approach is Libecap (1989), who, like Knight, emphasizes the role of distributional conflict. Different configurations of "property-rights" (rules which govern day-to-day interaction) entail different distributional consequences, and individuals and groups therefore engage in "contracting" to try to alter the rules for their own benefit, either privately among themselves or by lobbying the government. The contracting process by which the property-rights rules change is in turn governed by higher-level political rules. Exogenous parameter shifts are the basic cause of institutional change (p.16), and whether a parameter shift will lead to a change in the property-rights rules depends on the distribution of benefits both under the existing and proposed new systems, and on whether groups who expect to be losers from a change are able to block it under the rules which frame the political (rule-making) contest.

Libecap gives history an important role. Existing institutions influence the bargaining strength of different parties and often create groups with a vested interest in preserving the status quo. This can lead to a bias towards preserving the status quo, and more generally, makes institutional

[4] There are many possible reasons for differences in bargaining power; Knight focuses on differences in wealth.

change a path-dependent process: "Past political agreements on property institutions create the framework for responding to new common pool losses, the identities of the agents for and opponents of change, their effectiveness in political bargaining, and the range of feasible alternatives" (Libecap 1989, p.116). As a result, inefficient institutions can persist, and institutional change is usually incremental since it is often easier to achieve consensus on small adjustments than to effect major changes to existing rules.[5]

Ostrom (2005) uses a related but more complex approach involving a multi-layer nested hierarchy of rules. She distinguishes between "operational rules" which govern day-to-day interactions, "collective-choice rules" which are rules for choosing operational rules, and "constitutional rules" which are rules for choosing collective-choice rules. The process need not stop there; there may also be 'meta-constitutional rules", which are rules for choosing constitutional rules (the "rules" by which a civil war is fought, for example, might belong in this category). Ultimately, however, at the top of this pyramid, we arrive at a level at which there are no humanly-devised rules, but only a set of constraints which reflect the physical possibilities available to the players: in other words, Hobbes's "state of nature" (Ostrom 2005, p.58, p.211).[6,7]

As for the causes of institutional change, Ostrom recognizes both exogenous causes (such as technological change) and endogenous causes (such as the depletion of a resource over time). Following an impetus for institutional change, in order to analyze how rules are formed at one level, Ostrom temporarily treats the higher levels of rules as fixed (p.61). For example, constitutional and collective-choice rules are treated as exogenous when "operational rules" are being chosen. The process of institutional change is this: each individual calculates their expected costs and benefits from an institutional change, and if a "minimum coalition" necessary

[5] Knight (1995) also emphasizes the role of distributional conflict.
[6] Hodgson (2002) points out that even the state of nature as usually envisaged is not entirely institution-free: some minimal pre-existing institutions, such as language, must be implicitly assumed.
[7] Acemoglu and Robinson (2006) present a theory of constitutional change, in which two interest groups compete to implement their favored political system (the Rich favor dictatorship, the Poor favor democracy). They distinguish between *de facto* political power (the ability to use force) and *de jure* political power (control of political institutions). *De facto* power is transient, because collective action problems impede a group from using violence except at rare moments of crisis. In moments of crisis, the players find themselves, in effect, in a state of nature, and jostle for control of political institutions in order to convert their fleeting *de facto* power into durable *de jure* power. The model yields numerous predictions about the determinants of democracy as well as insight into the forces leading to democratic consolidation over time.

to effect change agrees to it, an institutional change can occur. What constitutes a "minimum coalition" is determined by the higher-level rules; for example, in a dictatorship the dictator alone might constitute a winning coalition; in a democracy, a majority would constitute a winning coalition. Therefore, whether an institutional change occurs ultimately depends on the higher-level rules and on how the decisionmakers perceive the likely effects of a change in rules.

Even if all players are fully rational, this process will not necessarily produce an efficient outcome. The players may try to design efficiency-enhancing institutions, but they may put their energy instead into trying to create institutions which will redistribute income. As long as the beneficiaries of institutional change cannot commit to compensate the losers, those with a vested interest in maintaining the status quo may be able to block beneficial change, or impose inefficient change.[8]

Inefficient outcomes can also result from the bounded rationality of the players (Ostrom, ch.4). They may hold different beliefs about the likely effects of a proposed change, and some or all of them may be wrong. Recognizing this, players may try to learn about the likely effects of a change both by experimenting with institutional innovations and by imitating successful institutions observed elsewhere. Thus, the overall pattern of institutional change may have an evolutionary flavor, and Ostrom's framework can accommodate institutional diversity in governing apparently similar transactional settings.

An important variant on the hierarchy-of-rules approach are theories which give "the state" (or elements thereof) a role as an actor with its own objectives, rather than viewing it primarily as a battleground in which groups compete to mold formal rules to their own advantage. In Kantor's (1998) framework, groups of constituents lobby politicians to change formal rules. The politicians may attempt to accommodate their demands, but also have their own objectives and face political and constitutional constraints. North (1981: ch.3) presents a model in which formal rules (property rights) are designed by a predatory ruler, whose objective is to maximize tax

[8] Levi (1990) emphasizes that formal rules can give "power" to certain groups, and that disadvantaged groups may try to force institutional change by "withdrawing their consent" from existing institutional arrangements.

revenue, rather than economic output or growth (although threats from potential rivals ensure that he cannot ignore efficiency considerations entirely).

The hierarchy-of-rules approach provides numerous insights, but still leaves several important questions unanswered. In particular, it has difficulty explaining why, in many cases, formal rules are ignored, or fail to produce their intended outcome. For example, Ostrom distinguishes between rules-in-form (rules which may be dead letters) and rules-in-use (rules which are actually followed) (p.138) but subsequently pays little attention to this distinction, as her scheme gives her no way to tell what rules will actually become rules-in-use.

One reason for this difficulty is that the hierarchy-of-rules approach is ill-equipped to deal with the role of informal rules in institutional change. Both Libecap and Ostrom acknowledge the importance of "informal rules", but this term is used rather vaguely, and an important class of informal rules are essentially ignored. To see this, let us distinguish three types of "informal rules".

First, the term "informal" is sometimes simply used to indicate that the rules are not written down, or are not enforced by the state. If such informal rules can be changed by agreement, then they fit easily into the hierarchy-of-rules framework. For example, a group of friends might make an informal agreement not to date each other's sisters, and to beat up anyone who defects. The corresponding collective-choice rule might, for example, be a rule of unanimity: the rule could be instituted or abolished by unanimous agreement. The informal agreements among fishermen described by Libecap (p.87), for example, fit this pattern.

Second, the term "informal rules" is sometimes used to refer to ethical codes or moral "norms", such as norms of fairness or reciprocity, which are internalized and directly reflected in players' preferences. These kinds of norms can impact institutional change by affecting the process of choosing among formal rules. For example, a norm of fairness can be modeled as directly reducing players' utility if an "unfair" outcome is observed (eg., Ostrom, p.122), and if players have such preferences, certain rules may not be adopted because they are perceived as unfair.

However, there is a third, important category of "informal rules" which are not internalized, but rather are followed because deviating from the rule is not individually rational if others follow it. Examples include "social norms" which use a multilateral reputation mechanism and a credible threat of punishment to generate trust among members of a community (see, eg., Kandori (1992), Greif (2006)), and "conventions", viewed as self-enforcing solutions to multiplayer coordination games (Sugden 1989). Of course, the informal "rules" may eventually come to be followed without rational evaluation, and socially experienced as moral or ideological rather than purely strategic constraints.[9] However, in general, if all others are following such rules, then even fully rational strategic players are also induced to follow them. This is crucial because, even though most people follow the norm without rational evaluation, if a rational mutant could achieve a higher payoff by deviating from the norm, given the evolutionary manner in which informal rules are usually held to emerge, it is unlikely that a norm could survive.

Because social norms and conventions generally evolve in a decentralized, "spontaneous" manner, they do not fit easily into the hierarchy-of-rules model of institutional change, which treats institutional change as the result of deliberate centralized rule-changing activity (though perhaps with unintentional outcomes). As a result, the hierarchy-of-rules models' treatment of formal and informal rules is asymmetric, and their treatment of informal rules is incomplete.

This may be a serious shortcoming because the evolution of informal rules is frequently an important part of the story of institutional change. For example, Lloyd's of London began as a coffee-house which became a focal gathering place for those involved in marine insurance in the early eighteenth century, and by the mid-nineteenth century it had evolved into a highly structured and formal marketplace for marine insurance and the dominant force in the world marine insurance industry (Kingston 2005). Much of this development occurred without any conscious efforts to design new formal rules or change old ones. For the first half-century of its existence Lloyd's had virtually no formal structure at all, and when a formal structure was eventually created, largely as a result of the impetus provided by the Napoleonic wars, formal rules were adopted mainly "to systematize a practice which had already been adopted to meet the

[9] As Cooter writes (1993, p.423): "The cooperative solution to a repeated game can be nothing more than the coincidence of self-interested strategies. Adherence to such strategies creates expectations, however, and expectations often turn into obligations. By this familiar but mysterious chemistry, strategies become social norms."

requirements of commerce as they arose." (Wright and Fayle 1928, p. 2). Even then, informal rules and reputation mechanisms remained the dominant mode by which participants at Lloyd's were constrained from opportunistic behavior. Because the business practices which evolved acquired the force of informal custom long *before* they were systematized as formal rules, the collective-choice process by which they became formalized is much less interesting than the spontaneous manner in which they initially arose.

Brousseau and Raynaud (2006) argue that many institutional arrangements begin as "private" local experiments, in which participation is voluntary, but that over time, through competition for adherents, economies of scale, and network effects, some (not necessarily optimal) institutions emerge as "winners", and become "solidified" as part of the institutional environment (higher-level institutions), so that participation in them becomes increasingly mandatory. Thus, in comparison to Ostrom's approach, in which institutional change involves a deliberate collective-choice process, Brousseau and Raynaud's approach envisions a more continuous and evolutionary kind of process by which lower-level institutions can, in a sense, gradually climb the hierarchical ladder to become more permanent, higher-level institutions. The development of Lloyd's appears to fit this pattern.

2.3 Formal Rules and Informal Rules (Williamson 2000; Roland 2004; North 1990)

As we have seen, the hierarchy-of-rules approach neglects the role of some kinds of informal rules; whereas the evolutionary theories neglect the role of collective action and the political process. The question therefore naturally arises as to how to integrate these theories: what is the interaction between these two mechanisms of institutional change, and under what circumstances will each apply? To a large extent, this turns on the interaction between formal rules, which are generally deliberately designed, and informal rules, which generally evolve spontaneously.

Williamson (2000) treats informal rules as providing the background within which formal institutions are "embedded". He distinguishes four "levels" of institutions, according to how quickly they change. At the "highest" level are the "institutions of embeddedness", including informal institutions, culture and norms, in which change occurs on the order of centuries or

millennia. At the second level, constrained by the institutions of embeddedness, are the high-level formal rules: constitutions, laws, and property rights. At this level, change takes decades or centuries. The third level is that of the "institutions of governance", at which the sets of rules ("governance structures") which govern day-to-day interactions ("contractual relations") are assumed to adjust so as to minimize transaction costs. Adjustment at the third level typically takes years. Finally, at the lowest level, the prices and quantities specified in individual contracts adjust continuously. Williamson recognizes the possibility of long-run feedbacks from lower to higher levels, but then consciously ignores it (2000: 596). Moreover, as defined by Williamson, "the NIE operates at two levels" (2000: 610), namely levels 2 and 3, while level 1 institutions are "an important but underdeveloped part of the story" (p.610). Thus, informal rules are effectively taken as exogenous and excluded from the scope of the analysis.

North (1990), in contrast, gives informal rules a central role in institutional change. In North's account, as in the hierarchy-of-rules approach, formal rules change as a result of deliberate (though boundedly rational) actions by organizations and individual entrepreneurs via a political process.[10] The impetus to try to change formal rules can come from exogenous or endogenous parameter changes (including learning). North identifies two kinds of exogenous parameter shifts: changes in relative prices (including technological change), and changes in preferences.

North combines this theory with a theory in which informal rules evolve alongside, and as extensions of, formal rules. Informal rules are reproduced though a cultural transmission process and evolve in an evolutionary manner. For North, informal rules include both internally enforced personal norms and third-party enforced social norms, as well as "extensions, elaborations and modifications of formal rules" (p.40), and are "part of the heritage that we call culture". Informal rules play a key role in institutional change because they change slowly and cannot be changed deliberately. Following a change of formal rules, the informal rules which "had gradually evolved as extensions of previous formal rules" (p. 91) survive the change, so that the result "tends to be a restructuring of the overall constraints - in both directions – to produce a new equilibrium that is far less revolutionary" (p. 91).

[10] North (1990, p.47) envisages a hierarchy with four levels of formal rules: constitutions, statute and common laws, specific bylaws, and individual contracts.

As a result, in North's view, institutional change is generally incremental rather than sudden, an accumulation of many small changes rather than occasional large changes. The process of institutional change is also path-dependent because individuals learn, and organizations develop, and ideologies form in the context of a particular set of formal and informal rules (1990, ch. 9). These organizations then may attempt to change the formal rules to their benefit, and over time this in turn may (indirectly) affect the informal rules. In general, there are multiple equilibria and no guarantee of an efficient outcome (p.80-81, p.136). Bounded rationality leading to unintended outcomes or inertia caused by pre-existing informal rules can prevent the implementation of an efficient institutional change. In effect, what TCE takes for granted is what North regards as the elusive objective: a system which rewards and replicates efficient institutional innovations.

Roland (2004) distinguishes between "fast-moving" (political) institutions (akin to formal rules), which can be changed quickly and deliberately via the centralized political process, and "slow-moving" (cultural) institutions (akin to informal rules), such as social norms, which change slowly because change is continuous, evolutionary and decentralized. In comparison with North, Roland gives informal rules a more central role as drivers of institutional change. He outlines his view of institutional change by analogy: tectonic pressures along fault lines (changes in slow-moving institutions) build up continuously but slowly, then suddenly provoke an "earthquake" that causes abrupt and substantial changes in fast-moving institutions (ie., formal rules).

2.4 The "Equilibrium View" of institutions (Calvert 1995, Aoki 2001, Greif 2006)

Despite significantly advancing our understanding of the role of informal rules in institutional change, North still concludes his book with the thought that "We need to know much more about [informal rules] and how they interact with formal rules" (1990, p.140). One strand of literature which has emerged in response to this challenge attempts to treat formal and informal rules within a unified framework by shifting the focus from the rules governing behavior to the behavior itself. Elements of this approach, which we will refer to as the "equilibrium view", can

be found in the work of Calvert (1995), Aoki (2001), Greif and Laitin (2004), and Greif (2006), although the exposition below is not intended to represent any of their views directly.

To begin with, imagine an "underlying" game (or "state of nature") in which there are no "humanly devised constraints" on behavior, so that all physically possible actions are potentially relevant. That is, only constraints which are truly exogenous (at least in the short run), such as the laws of physics, resource endowments, technology, capital stock, and so on, constrain the players. There are generally many possible equilibria in the underlying game – in other words, institutions, behavior and outcomes cannot be deduced solely from technological constraints.

In essence, the equilibrium view of institutions regards the essential role of both formal and informal "rules" as being devices to enable players to coordinate on one of these many equilibria by helping them to achieve a shared set of beliefs about each other's behavior both on and off the path of play; and it is these beliefs, rather than the rules, which are fundamental. Thus, Calvert (1995, pp.22-23) writes: "There is, strictly speaking, no separate animal that we can identify as an institution. There is only rational behavior, conditioned on expectations about the behavior and reactions of others… "Institution" is just a name we give to certain parts of certain kinds of equilibria." Aoki (2001) defines institutions as stable, shared systems of beliefs about the expected behavior of the members of a society in various contingencies. Greif (2006: 30) adopts a more expansive definition: "An institution is a system of rules, beliefs, norms and organizations that together generate a regularity of (social) behavior".

While the exact definition of institutions differs among these authors, the common thread is that institutions are identified with these equilibrium patterns of behavior rather than the rules which induce the behavior. In equilibrium, each agent is constrained both by the exogenous physical constraints in the underlying game, and also by the endogenous institutional "rules of the game", which reflect the strategies of the other players[11] (possibly summarized in the form of formal and informal "rules"). In maximizing their welfare subject to these constraints, agents choose strategies which, in the aggregate, and perhaps unintentionally, give rise to expectations which

[11] It is not necessary for players to know the details of all the other player's strategies. It is sufficient that, from experience or otherwise, they can accurately predict what payoffs they can obtain by taking various feasible actions (Aoki 2001).

reinforce the constraints on everyone else. Thus, institutions emerge as endogenous equilibrium outcomes, reflecting a socially constructed "reality".

For example, informal social norms enforced within a community can be viewed as coordinating the beliefs of many (rational) players about the actions that each will take on and off the path of play. Similarly, a formal rule making one player a "judge" does not change that player's set of physically feasible actions, but it *may* systematically alter people's perceptions about how those actions are to be interpreted: by virtue of her role, the judge can give orders which would not be followed if she were considered an ordinary citizen.

In the Equilibrium View, institutional change becomes fundamentally not about changing rules, but about changing expectations. For example, a formal rule "forbidding" some kind of behavior does not change the set of physically feasible actions available to any player, so the key issue is how it affects people's expectations about the consequences of those actions. A rule which fails to shift people's expectations in the desired way may have no effect at all (Aoki 2001: 231): in Ostrom's terms, it will be a "rule-in-form" but not a "rule-in-use". But more generally, it might have an unintended effect. For example, in one country a new law granting power to policemen to detain travelers at airports might be expected to enable the police to prevent terrorism, while in another it might be generally expected to enable the police to extract bribes by threatening to make people miss their flights. The hierarchy-of-rules approach obscures these possibilities because it considers the enforcement of rules separately from their content; in the equilibrium view, enforcement is endogenous.

Both deliberate, centralized and evolutionary, decentralized institutional change are compatible with the equilibrium view. Exogenous parameter shifts such as changes in technology or preferences can disrupt an equilibrium, leading individuals and organizations to try to change the "formal rules" in order to achieve a coordinated shift of many players beliefs about each others' strategies. Alternatively, gradual changes in parameters might cause gradual adjustments to expectations and behavior. Since the formal rules remain unchanged, this kind of institutional change would be manifested as changing "informal rules".

Greif and Laitin (2004) highlight the potential for endogenous sources of change. They use the term "quasi-parameters" to refer to parameters which are exogenous in the short run, but which gradually change as a result of the play of the game, such as the income distribution, or the information available to the players. Changes in quasi-parameters may either broaden the range of situations in which the existing pattern of behavior (institution) is an equilibrium, or may undermine the existing institution, leading to an "institutional disequilibrium" and an impetus for institutional change.

Greif and Laitin argue that institutional change will follow a "punctuated equilibrium" process in which gradual changes in quasi-parameters lead to a "crisis" (and deliberate institutional change) when it becomes clear that existing patterns of behavior no longer constitute an equilibrium. Aoki (2001: 243) argues that institutional change frequently involves short, turbulent periods of deliberate institutional change and experimentation, interspersed with longer periods during which these experiments are weeded out through competition.

2.5 Mental Models (North 2004, Aoki 2006)

In all of the theories surveyed above, institutions ultimately arise as a result of purposeful human problem-solving as they attempt, in a boundedly-rational way, to process incomplete information (about the environment and the strategies of other players) in a complex and changing environment. Therefore, the way people process information and solve problems may have an impact on the nature of institutional change. For example, people may systematically misperceive opportunities in a new environment; or some equilibria may become more focal than others because of the way in which people learn (individually and collectively).

The literature on "behavioral economics" provides substantial empirical support to the idea that people are boundedly rational, and according to Williamson (2000: 600) "there is close to unanimity" on the idea of bounded rationality in institutional economics. Abandoning the rational actor model in favor of more realistic views of human cognition is an enormously complex undertaking, however, and so far the links between bounded rationality at the level of the individual and institutional change at the aggregate level remain unclear.

For some authors, bounded rationality, though a realistic view of people's cognitive abilities, may not be an essential component for understanding institutions or institutional change. In TCE, as we have seen, bounded rationality helps to explain the existence of institutions, but plays no direct role in institutional change, because efficient institutions are assumed to emerge. Also, there is nothing inherently non-rational about using shared knowledge of past institutions as a coordinating device in novel situations (Sugden 1989).

Other theories give bounded rationality a potential role in institutional change. In the hierarchy-of-rules view (eg., Ostrom), bounded rationality can affect the direction of institutional change because actors may have an incorrect understanding of the effects of different rule changes. But does this just add "noise" to the system, or does it have a systematic and predictable effect on institutional change, and under what circumstances?

In North's (2004) framework, economic actors have "mental models" which reflect their understanding of the world and which they use to evaluate the desirability of particular rule changes. Over time, as they learn about the world, they revise their mental models and may alter their perceptions of the effects of alternative rules and of the set of possible alternative rules. These changed perceptions may lead them to attempt to change rules for their own benefit, and this activity provides the impetus for institutional change. Since existing institutions provide incentives to create particular kinds of organizations and to invest in particular kinds of skills and knowledge, and because this learning and organization-building in turn affects people's perceptions of new possibilities, past institutions exert a substantial influence on the direction of institutional change.

North views institutions as rules. For Aoki (2006), in contrast, an institution as "a summary representation of the salient features of a moving equilibrium path" (what we have referred to as the "Equilibrium View"), so that institutional change involves a movement to a new equilibrium path. However, because of bounded rationality, each individual agent has an incomplete understanding of this overall equilibrium, and instead observes only a truncated, simplified version (or mental model) of the overall game as it applies to him or her – what Aoki calls an

"internalized game-form". From this perspective, "institutions" can be viewed not as a set of rules, but as a set of shared perceptions which tie all the individual, truncated games together. Exogenous or endogenous parameter shifts can break down the existing overall equilibrium, so that some agents perceive that their strategies are no longer optimal, without necessarily understanding why. In the ensuing "institutional crisis", agents individually experiment with new strategies as they collectively grope towards a new overall equilibrium. Eventually, their strategies and mental models will become mutually consistent within a new equilibrium, either via a decentralized process (in which case, focal points may play an important role in equilibrium selection), or possibly through the leadership of a politician or organization (in which case, because of bounded rationality, the outcome may still be unintended).

These theories suggest that a key to understanding institutional change is an understanding of how people learn and revise their "mental models". For example, North (chapter 3) argues that people often reason by analogy. Overall, ongoing research in cognitive psychology and behavioral economics offers the promise of enabling a deeper understanding of many aspects of institutional change (for an optimistic assessment, see Ostrom 2005, ch. 4), although research in this area is still at an early stage.

3. Institutional Inertia and the Role of History

Institutions, by definition, are relatively durable rules (or in the equilibrium view, stable sets of expectations giving rise to a regularity of behavior). Either exogenous or endogenous parameter changes, including cognitive developments, can lead to institutional change. Yet, not all parameter changes lead to institutional change. For example, prices may change but the overall mode of functioning of the market is not affected. Therefore, perhaps as important as understanding the causes of institutional change is understanding why change often does not occur – ie., the reasons for institutional inertia. Institutional inertia hinders efforts at institutional change, impedes inter-society transplants of institutions, and makes institutional change "overwhelmingly incremental" (North 1990, p.89). Several sources of inertia are identified in the works discussed above.

An important source of inertia is the combination of bounded rationality and risk-aversion. Given the uncertainty associated with new institutions, boundedly rational people may be unwilling to experiment, especially with radical changes.

Institutional inertia can also result from free-rider problems which impede collective action to change formal rules. For example, voting, protesting, joining political associations, and even learning about the impact of potential policies may all be individually non-rational actions, even if the individual cares deeply about the result. North (1981: ch. 5) notes that the free-rider problem is an important source of institutional stability, and stresses the role of ideology in overcoming this free-rider problem.

North (1990) regards informal rules as the major source of institutional inertia. In his view, informal rules which evolved as extensions of previous configurations of formal rules persist after a change in formal rules, making discontinuous change a relative rarity. In the equilibrium view, inertia can result from the difficulty in achieving a coordinated shift of expectations. In particular, Aoki (2001) argues that because of strategic linkages and complementarities between institutions in different "domains" in the economy – for example, between particular kinds of political systems, labor market institutions, and mechanisms of corporate governance - institutional change in one domain can have knock-on effects on institutions in other domains. Conversely, radical change is unlikely to occur in a single domain in isolation.

Institutional inertia is also closely related to the role of history. All the theories surveyed above, except TCE, agree that institutional change is a path-dependent process: "the consequence of small events and chance circumstances can determine solutions that, once they prevail, lead one to a particular path" (North 1990: 94). The past can constrain the future in many ways. First, as Libecap emphasizes, existing institutions can affect the configuration of interest groups and their bargaining power, and groups with a vested interest in the status quo may attempt to block subsequent institutional change. More generally, capital of all kinds – buildings, machines, human capital, social capital, and technological and organizational knowledge – accumulated under one set of institutions can gradually alter the set of technologically feasible institutions and thereby affect future institutional development.

In addition, previous institutions provide focal points which can affect equilibrium selection in novel situations. In terms of the equilibrium view, coordinating on one of the many possible institutional equilibria in society is a giant society-wide coordination game, and as a result, there is no guarantee that efficient institutions will emerge even in a static sense. Each actor will rationally follow strategies which constitute a pareto-inferior equilibrium, if she expects the others to do so. Even if the actors collectively perceive the potential for better institutions, engineering the coordinated shift of expectations necessary to move to a better equilibrium may prove problematic.

Greif and Laitin (1994) generalize these arguments: knowledge, resource ownership, income distribution, and other "quasi-parameters" can all be affected by past institutions, and affect both the future institutional choice set and the choice of institutions within that set.

Yet this does not imply that relatively efficient institutions will *not* emerge. In some situations, the assumption that competition will weed out inefficient institutions may be justifiable. Boundedly rational actors may sometimes misperceive the effects of institutional changes, but it would be remarkable if they were *always* wrong. Transaction costs may sometimes hinder efficient bargains, they need not always do so, especially if a bargain exists which can make (almost) all of the players better off. Ellickson (1991, ch.10) argues that even in the case of informal norms which evolve spontaneously rather than being subject to deliberate design, there may be a tendency for efficient norms to emerge. The question of whether and when efficient institutions will emerge is ultimately an empirical one, and the answer may well depend on the context.

4. Conclusion.

The appropriate model for studying institutional change may be largely a matter of context. For situations in which competition will weed out inefficient institutions, institutional change is of relatively minor interest and the Transaction Cost view is likely to be appropriate. In micro-level situations in which changes in formal rules occur within a stable political context, and have

relatively predictable effects on behavior, treating institutional change as an outcome of a political process, as in the hierarchy-of-rules approaches, has proven useful in many real-world settings. It is incomplete, however, because it cannot explain why some formal rules become effective and others do not. The equilibrium view of institutions provides a more complete theory by treating informal and formal rules within an integrated framework, and is therefore useful as a broad conceptual framework for understanding institutional change, particularly in cases in which changes in formal rules may fail to have their desired effects. However, it may introduce unnecessary complexity in the many real-world cases in which formal rules are relatively straightforward and effectively enforced. Finally, all of these theories can benefit from further theoretical and empirical work to clarify the role of cognition and bounded rationality in institutional change.

References

Acemoglu, Daron, and James Robinson (2006): *Economic origins of Dictatorship and Democracy*. Cambridge University Press.

Alchian, Armen (1950): "Uncertainty, Evolution and Economic Theory", *Journal of Political Economy*, 58(3).

Aoki, Masahiko (2001): *Towards a Comparative Institutional Analysis*. MIT press.

Aoki, Masahiko (2006): "Mechanisms of Endogenous Institutional Change", SIEPR Discussion Paper No. 05-13.

Brousseau, Eric and Raynaud, Emmanuel (2006): "The Economics of Private Institutions: An Introduction to the Dynamics of Institutional Frameworks and to the Analysis of Multilevel Multi-Type Governance". Downloaded from SSRN: http://ssrn.com/abstract=920225

Calvert, Randall (1995). "Rational Actors, Equilibrium and Social Institutions", in Jack Knight and Itai Sened, editors, *Explaining Social Institutions*. University of Michigan press.

Campbell, John L. (2004). *Institutional Change and Globalization*. Princeton University Press.

Cooter, Robert (1993). "Against Legal Centrism". *California Law Review* 81.

Ellickson, Robert (1991). *Order without law*. Harvard University Press

Greif, Avner and David Laitin (2004): "A Theory of Endogenous Institutional Change", *American Political Science Review* 98(4).

Greif, Avner (2006): *Institutions and the Path to the Modern Economy.* Cambridge University Press.

Hodgson, Geoffrey (2002). "The Evolution of Institutions: An Agenda for future Theoretical Research", *Constitutional Political Economy* 13.

Kandori (1992). "Social Norms and Community Enforcement". *Review of Economic Studies* 59.

Kantor, Shawn Everett (1998). *Politics and Property Rights: The closing of the open range in the postbellum south.* Chicago, IL: University of Chicago press.

Kingston (2005). "Marine insurance in Britain and America, 1720-1844: A Comparative Institutional Analysis". Mimeo, Amherst College, November 2005.

Knight, Jack (1995): "Models, Interpretations and Theories: Constructing Explanations of Institutional Emergence and Change" in Jack Knight and Itai Sened, editors, *Explaining Social Institutions*. University of Michigan press.

Levi, Margaret (1990). "A logic of institutional change", in Karen Schweers and Margaret Levi, editors, *The Limits of Rationality*. University of Chicago Press.

Libecap, Gary D. (1989). *Contracting for Property Rights.* Cambridge University Press.

North, Douglass (1981): *Structure and Change in Economic History*. WW Norton & Co.

North, Douglass (1990): *Institutions, Institutional Change and Economic Performance.* Cambridge University Press.

North, Douglass (2005): *Understanding the process of economic change*. Princeton University Press.

Ostrom, Elinor (2005). *Understanding Institutional Diversity*. Princeton University Press.

Putnam, Robert (1993). *Making Democracy Work*. Princeton University Press.

Roland, Gerard, "Understanding Insitutional Change: Fast-moving and Slow-moving institutions". *Studies in Comparative International Development* 38(4), Winter 2004, pp.109-131.

Sugden, Robert (1989): "Spontaneous Order", *Journal of Economic Perspectives* 3(4).

Thelen, Kathleen (2004): *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States, and Japan.* Cambridge University Press.

Williamson (2000): "The New Institutional Economics: Taking Stock, Looking Ahead". Journal of Economic Literature 38, Sept 2000, pp. 595-613.

Wright, Charles, and C. Ernest Fayle (1928). *A History of Lloyd's*. London: MacMillan and Company.