

# Comparing theories of institutional change

CHRISTOPHER KINGSTON\*

*Amherst College, Amherst, Massachusetts, USA*

GONZALO CABALLERO

*University of Vigo, Vigo, Spain*

**Abstract:** This article compares a variety of theoretical approaches to conceptualizing institutional change. Our goal is neither to discover the ‘best’ theory, nor to attempt to build a new one. Rather, we wish to compare how the theories we consider agree or differ with respect to the causes, process, and outcomes of institutional change. Some of the theories we discuss emphasize the deliberate creation of institutions through the political process, while others emphasize the spontaneous emergence of institutions through evolutionary processes. Still others combine elements 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 habit, learning, and bounded rationality. We also consider theoretical explanations for institutional inertia and path-dependence.

## 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 creates a dilemma for scholars undertaking empirical studies of institutional change, as they consider what theoretical puzzles they should be trying to address. If the empirical study is to inform theory, and be informed by theory, then one can either tie oneself to a particular theoretical approach and interpret one’s empirical findings in the light of that approach, or one can attempt to link one’s work to a wider variety of theories. The first approach, however, tends to lead to a balkanization of the literature, while the

\*Email: [cgkingston@amherst.edu](mailto:cgkingston@amherst.edu)

We thank Veneta Andonova, the editor, and five anonymous referees for detailed and helpful comments and suggestions.

second inevitably leads to some mismatch in the conceptual tools used and makes it difficult to engage with theory in a satisfactory way.

This paper grew out of our confrontation with these dilemmas when undertaking our own empirical studies of institutional change (Kingston, 2007; Caballero, 2008). The sometimes ambiguous and inconsistent use of terminology in the literature, even regarding such commonly used terms as ‘norms’, ‘transactions costs’, ‘formal’, or ‘informal’ rules, and the meaning of the term ‘institutions’ itself, has the potential to create confusion and make it hard to compare different theories. Our goal here is not to advance new definitions of these terms, or to choose between the existing definitions, but simply to try to map out some of the areas of confusion within a common framework, as a necessary first step towards achieving consensus or at least understanding. Similarly, in surveying theories of institutional change, it is not our goal to discover the ‘best’ theory, nor do we attempt to build a new one. Rather, we compare and contrast existing theories in an effort to draw out some of the commonalities and differences in a way we hope may help to promote dialogue and reduce confusion among theorists of institutional change, and may also prove useful to those carrying out empirical studies of institutional change by highlighting the main possibilities that an empirical researcher might consider. The theories we survey were developed to explore a variety of settings, so they naturally reflect different perspectives on many of these questions. Evaluating the appropriate range of contexts within which to apply the various theories is ultimately a task for empirical work.

Among the questions we wish to consider are:

- Under what circumstances is institutional change best viewed as a spontaneous, evolutionary process rather than as the outcome of deliberate design? How can these two views be reconciled and combined?
- How should we think about the interaction between formal rules and informal constraints?
- What determines the pace of institutional change?
- What is the role of politics and collective action?
- What is the role of ‘bounded rationality’?
- What roles do exogenous shocks and endogenous parameter changes play as drivers of institutional change? What kinds of endogenous processes are involved?
- When do institutions fail to change, and why? What are the main sources of inertia that make institutions relatively durable over time?
- Under what circumstances, and what time horizons, will efficient institutions tend to emerge?
- What kinds of equilibria are likely to be selected when there are multiple equilibria?
- When, and how, does history matter?

Given the volume of relevant literature, our survey is necessarily somewhat selective. We have sought to include what we perceive as the main strands of

thought in institutional economics, but for example, we do not discuss historical institutionalism in sociology and political science (Thelen, 2004; Campbell, 2004), although many related issues arise in that literature.

To organize our discussion, we identify two broad categories of processes of institutional change. We begin (Section 2) by considering theories in which institutions are purposefully designed and implemented in a centralized way, either by a single individual (such as, for example, when a king issues a royal decree, or the manager of a firm implements a new organizational structure) or by many individuals or groups interacting through some kind of collective choice or political process in which they lobby, bargain, vote, or otherwise compete to try to implement institutional changes which they perceive as beneficial to themselves, or to block those they view as undesirable.

Second, we consider theories of ‘evolutionary’ institutional change (Section 3), in which new institutional forms periodically emerge (either at random or through deliberate design) and undergo some kind of decentralized selection process as they compete against alternative institutions. Those institutions that prove successful at surviving the selection process spread, by imitation or replication, while unsuccessful institutions die out. As a result, overall institutional change occurs ‘spontaneously’, through the uncoordinated choices of many agents, rather than in a centralized and coordinated manner.

Both the evolutionary and design-based approaches to understanding institutional change are useful in particular settings. In many real-world processes of institutional change, however, both unintentional, evolutionary processes and intentional processes of design are at work, and it will often be difficult to cleanly separate the two. For example, the impetus for attempts at deliberate institutional design may reflect gradual underlying changes in parameters, beliefs, or knowledge, which result from the spontaneous evolution of the existing institutions over time; and even when deliberate attempts are made to design and implement new institutions, competition or other evolutionary processes may subsequently play a role in determining which kinds of institutions survive and spread. The main source of variation (mutation) driving an evolutionary process of institutional change will often be the deliberate attempts of the actors to craft new rules; and the selection process through which some institutions are ultimately selected may also be driven by the deliberate choices of actors to adopt one or another of the competing alternatives, thereby playing an unintentional role in a collective process of institutional ‘design’. Therefore, building theories in which processes of evolution and design are integrated within a broader framework is a priority for current and future research, and in Section 4 we outline some theories which integrate aspects of these two perspectives, giving particular emphasis to what we refer to as the ‘Equilibrium View’ of institutions (Section 5).

While we are primarily interested in change, institutions also exhibit considerable inertia (resistance to change, or ‘lock in’). We explore sources of institutional inertia in Section 6, as well as the important roles of bounded

rationality and endogenous preferences in both inertia and change; and we also consider a variety of arguments for why institutional change is often a ‘path-dependent’, historical process in which initial conditions and stochastic events can have lasting effects. Section 7 concludes.

### *Definitional issues*

Before we can discuss institutional change, we must define what we mean by ‘institutions’. Unfortunately, the appropriate definition is far from a settled issue. The theories we will consider below adopt a variety of definitions, and because different authors use different definitions of institutions, this naturally influences their views of institutional change. There is a considerable literature dealing with the question of how best to define institutions (see, for example, Hodgson, 2006).

The most commonly cited definition is that advanced by Douglass North: institutions ‘are the rules of the game in a society, or, more formally, are the humanly devised constraints that shape human interaction’ (North, 1990: 3). Institutions ‘reduce uncertainty by providing a structure to everyday life’ (*ibid.*), and include both formal rules such as laws and constitutions, and informal constraints such as conventions and norms.<sup>1</sup> Many authors adopt some variant of North’s definition: fundamentally, institutions are viewed as durable rules which govern human interactions, and which are also ‘humanly devised’ (so, for example, technological constraints like the ‘laws’ of physics are not institutions).

Most authors distinguish between ‘formal’ and ‘informal’ rules or constraints, but this distinction is drawn in a variety of ways. The term ‘formal’ is often taken to mean that the rules are made explicit or written down, particularly if they are enforced by the state, whereas informal rules are implicit; another interpretation is that formal rules are enforced by actors with specialized roles (Milgrom *et al.*, 1990), whereas informal codes of behavior are enforced endogenously by the members of the relevant group. ‘Informal constraints’, as North notes, ‘defy, for the most part, neat specification’ (*ibid.*: 36), but include ‘codes of conduct, norms of behavior, and conventions’ as well as ‘extensions, elaborations and modifications of formal rules’, and ‘are a part of the heritage that we call culture’ (*ibid.*: 36–40).

<sup>1</sup> North (1990: 4) makes what he regards as ‘a crucial distinction’ between organizations (such as firms, universities, or political parties) and institutions. For the purpose of analyzing overall societal change, North implicitly treats organizations as entities within which collective action and agency problems have been solved, so that he can abstract away from their internal governance and decision-making processes, and treat them as unitary actors. Thus for North, organizations are ‘players’ of the game, and as they pursue their objectives, they act as agents of institutional change. Many other authors, however, study the formal and informal mechanisms of internal institutional governance within organizations, and North himself does not deny that organizations are themselves also institutions; see Hodgson (2006) for correspondence in which North clarifies this point. In this paper, we are concerned both with theories which can shed light on changes in the internal governance of organizations and with theories dealing with overall societal change.

But while social norms enforced by the community – conventions which help players to coordinate their behavior and expectations – and internalized ethical codes such as those reflecting religious beliefs can all be viewed as informal constraints, they are distinct phenomena which may involve different mechanisms of change, and may have different short- and long-run effects on the broader pattern of institutional change. This can create confusion when trying to compare how different theories of institutional change deal with the role of informal constraints. For example, some authors have regarded informal constraints as essentially immutable, or taking centuries to change, but others have noted that at least some kinds of informal constraints can change quickly. We return to this issue at various points throughout the paper.

Because we want to be able to compare different theories, we will not commit to any particular definition of institutions, but will discuss a number of alternative definitions in the course of the paper. For the moment, however, let us treat institutions loosely as the (formal and informal) ‘rules of the game’ in a society. Then the question of institutional change becomes: how do the rules change?

## 2. Collective-choice theories of institutional change

Many authors treat institutional change as a centralized, collective-choice process in which rules are explicitly specified by a collective political entity, such as the community or the state, and individuals and organizations engage in collective action, conflict, and bargaining to try to change these rules for their own benefit.

In this vein, Libecap (1989) explores the origins of ‘property rights’ rules (viewed as rules which govern day-to-day interaction). Different configurations of property rights entail different distributional consequences, and individuals and groups therefore engage in bargaining, lobbying, and political action to try to alter the rules for their own benefit. Libecap refers to this rule-changing activity as ‘contracting’. Contracting is itself a game governed by a higher level of political rules, and these higher-level rules, together with the activities and perceptions of the actors therefore shape the direction of institutional change of the lower-level (property rights) rules.

Ostrom (2005) uses a related approach involving a multi-layer nested hierarchy of rules. She distinguishes between ‘operational rules’ (rules that govern day-to-day interactions), ‘collective-choice rules’ (rules for choosing operational rules), and ‘constitutional rules’ (rules for choosing collective-choice rules). 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).<sup>2</sup>

<sup>2</sup> 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: 58, 211). Hodgson (2004: 20) points out,

In order to analyze how rules are formed at one level, Ostrom temporarily treats the higher levels of rules as fixed (*ibid.*: 61). For example, when ‘operational rules’ are being chosen, constitutional and collective-choice rules are treated for the moment as exogenous. 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 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, as in Libecap’s theory, whether an institutional change occurs ultimately depends on the higher-level rules and on how the decision makers perceive the likely effects of a change in rules.

What causes institutional change? Libecap regards exogenous parameter shifts as the basic source of impetus for institutional change (Libecap, 1989: 16). 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 that frame the political (rule-making) contest. Ostrom recognizes both exogenous causes of institutional change (such as technological change) and endogenous causes (such as the depletion of a resource over time).

Libecap sees institutional change as a ‘path-dependent’ process.<sup>3</sup> History matters because ‘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’ (*ibid.*: 116). Frequently, existing institutions create groups with a vested interest in preserving the status quo, which can impede institutional change and enable inefficient institutions to persist. Furthermore, even when these impediments are overcome, institutional change is usually incremental since it is often easier to achieve consensus on small adjustments than to effect major changes to existing rules.

Similarly, Ostrom argues that if the beneficiaries of institutional change cannot commit to compensate the losers, powerful groups may be able to block beneficial change or impose inefficient change. One important reason why Coasean

however, 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.

<sup>3</sup> The term ‘path-dependent’ refers to stochastic processes with dynamic properties which do not guarantee convergence to a unique, globally stable equilibrium (David, 2007). In the context of institutional change, it implies that the institutions observed at any point in time may in part be a function not just of current technology, but also of precedent institutions and technologies. This may be because, for example, shared historical experiences under one set of institutions help people to form mutually consistent expectations and thereby coordinate their behavior in the future; or because of various kinds of sunk costs which would make reorganization costly (David, 1994). A variety of other ways in which ‘history matters’ are discussed below in Section 6.

bargaining would not necessarily clear the path for efficient institutional change is the lack of an external authority to enforce intertemporal bargains.<sup>4</sup> A further impediment to efficient institutional change, highlighted by Ostrom (2005: chapter 4) is the bounded rationality of the players: some or all of the players may hold incorrect beliefs about the likely effects of a proposed institutional change. Furthermore, she argues that because they recognize their own bounded rationality, the players may experiment with institutional innovations and attempt to imitate successful institutions observed elsewhere. Thus, although change is deliberate, the overall pattern of institutional change may have an evolutionary flavor, and like Libecap's, Ostrom's framework can accommodate institutional diversity in governing apparently similar transactional settings.

Many theories give 'the state', or political actors such as judges and politicians, an independent role in the rule-making process. Alston (1996: 26–27) outlines the general framework: 'Institutional change can be thought of as the result of supply and demand forces in a society. We can think of demanders as constituents and suppliers as the government . . . Institutional change results from the bargaining actions of demanders and suppliers.' Here, we might imagine a continuum of possible theories, depending on the degree of autonomy attributed to political actors. On one extreme, political actors can be viewed as simply reflecting the interests of particular groups, so that their individual interests are relatively unimportant and the political process remains essentially a battleground in which interest groups compete to mold formal rules to their own advantage.

Other theories, however, give political actors a more independent role. For example, in Kantor's (1998) framework, groups of constituents lobby politicians to change formal rules, and the politicians have incentives to be responsive to their constituents' demands. However, the politicians also have their own objectives and face other political and constitutional constraints.

In the institutional economics of John R. Commons, the judiciary plays the key role in mediating the creation of new rules. Commons focused on 'how the politically determined rules of the game affect economic behaviour and performance' (Kaufman, 2007: 16), where the rules are created and enforced by various groups and organizations, but particularly by the state. Commons argued that if existing rules became unsuitable (a 'limiting factor'), then individuals or

<sup>4</sup> Acemoglu and Robinson (2006) present a theory of constitutional change which highlights this dilemma. In their theory, disenfranchised groups can use violence to bring about constitutional change, but, because of collective action problems, they can only organize violence at rare (and exogenous) moments of crisis. Violence is destructive, so at moments of crisis there is an opportunity for a mutually beneficial bargain between disenfranchised groups and the incumbent ruling groups: the ruling groups would agree to carry out reforms, in exchange for which the disenfranchised groups would refrain from carrying out a revolution. However, because the disenfranchised groups' opportunity to use force is fleeting, the ruling groups cannot credibly commit themselves to honor their commitments to reform after the moment of crisis has passed. Therefore, a revolution, though destructive and costly, may be the only way for currently disenfranchised groups to credibly constrain the policy choices of future governments.

groups would attempt to change them through the courts or by legislation. The courts therefore play a key role in determining the direction of institutional change: a process Commons called ‘artificial selection’ to differentiate it from evolutionary theories involving ‘natural selection’ (Hodgson, 2004: chapter 13).

Finally, at the other extreme, some theories give political actors (or ‘rulers’) a central role as autonomous drivers of institutional change. For example, the view of the state as a benevolent, autonomous dictator, which is implicit in much of standard neoclassical economics, could be placed in this category. North (1981: chapter 3) presents a model in which formal rules (property rights) are designed by a predatory ruler whose objective is to maximize tax revenue, rather than economic output or growth. The interests and objectives of the ruler’s subjects are of secondary importance (although threats from potential rivals ensure that the ruler cannot ignore efficiency considerations entirely).

Which of these perspectives on the role of political actors is most appropriate? The answer would appear to depend on the political structure in a given context: in Ostrom’s terms, it will depend on the content of the higher-level rules. These in turn may be influenced by other social and economic variables, and by history.<sup>5</sup>

Although they provide numerous insights, theories which view institutional change as the outcome of a deliberate, collective-choice process of rule-creation also leave several important questions unanswered. In particular, they have difficulty explaining why, in many cases, formal rules are ignored, or fail to produce their intended outcome. For example, Ostrom (2005: 138) distinguishes between rules-in-form (dead letters) and rules-in-use (rules which are actually followed), but subsequently pays little attention to this distinction, as her conceptual framework gives her no way to tell what rules will actually become rules-in-use.

One reason for this difficulty is that the collective-choice approach is ill-equipped to deal with some kinds of ‘informal rules’. Both Libecap and Ostrom acknowledge the importance of informal rules, but this term is used rather vaguely, and an important class of informal rules is 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 collective-choice 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

<sup>5</sup> Thus, for example, Acemoglu and Robinson (2005) argue that after 1500, some European countries experienced a substantial growth in Atlantic trade which increased the political power of merchant groups in those countries. The growing strength of merchant groups constrained the power of monarchs, and, as a result, these countries developed institutions which were more conducive to economic growth than countries with lower levels of trade.



be instituted or abolished by unanimous agreement. The informal agreements among fishermen described by Libecap (1989: 87), for example, fit this pattern.

Second, the term 'informal rules' is sometimes used to refer to ethical codes or moral 'norms' 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 (Ostrom, 2005: 122), and if players have such preferences, certain rules may not be adopted because they are perceived as unfair.

However, there is an important third category of 'informal rules', which are not deliberately designed, but are nevertheless 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, for example, Kandori, 1992; Greif, 2006), and 'conventions', viewed as self-enforcing solutions to multiplayer coordination games (Sugden, 1989). Of course, these informal 'rules' may eventually come to be followed without rational evaluation, and socially experienced as moral, ideological, or 'cultural' rather than purely strategic constraints.<sup>6</sup> However, if all others are following such rules, then even fully rational strategic players may also be induced to follow them. This is crucial because, even if most people follow the norm without rational evaluation, it is unlikely that a norm could evolve or survive if a rational mutant could achieve a higher payoff by deviating from it.

This third category of informal rules (social norms and conventions) do, sometimes, change over time, but they generally evolve in a decentralized, 'spontaneous' manner, so they do not fit easily into the collective-choice models.<sup>7</sup> This may be a serious shortcoming in some contexts because the evolution of informal rules is frequently an important part of the story of institutional change. For example, in the early 1700s, Lloyd's coffeehouse became a gathering place for individual merchants and underwriters transacting marine insurance in London. By the mid-1800s, Lloyd's had evolved into a highly structured marketplace for marine insurance and the dominant force in the world marine insurance industry (Kingston, 2007). Much of this institutional development occurred without any conscious efforts to design new formal rules or change old ones. Indeed, for the first half-century of its existence, Lloyd's had virtually no formal structure at all,

<sup>6</sup> As Cooter writes (1993: 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.'

<sup>7</sup> Hodgson (2004) criticizes the work of John Commons on similar grounds, arguing that by viewing institutions as the outcome of subordination to a 'single will', such as a powerful individual or group, and by focusing on changes in legal institutions, Commons neglected the possibility of self-organizing or 'spontaneous' processes of institutional emergence and change.

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 requirements of commerce as they arose’ (Wright and Fayle, 1928: 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, to understand the origins of these practices it is necessary to study the spontaneous and evolutionary manner in which they initially arose as well as the collective-choice process by which they later became formalized.

### 3. Evolutionary theories of institutional change

A large body of literature treats institutional change as an evolutionary process. Here, we use the term ‘evolutionary’ primarily to refer to processes which satisfy the core Darwinian evolutionary principles of variation (a source of mutations), selection (survival of ‘successful’ traits), and inheritance (a process by which the successful traits are replicated). Note that this does not rule out rational action. Indeed, in most accounts of evolutionary institutional change, the ultimate impetus for institutional change comes from deliberate human actions (including learning, imitation, and experimentation). The key difference between evolutionary theories and the theories discussed in the previous section has to do with the selection process which determines which rules ultimately emerge. In the evolutionary theories, there is no central mechanism (such as legislation) which causes a coordinated shift in the rules perceived by all the players, or in their behavior or beliefs. Instead, new rules or behaviors (mutations), which may be either randomly or deliberately generated, undergo some kind of decentralized selection process, as a consequence of which some (successful) institutions spread through the population, while other (unsuccessful) institutions die out. Thus, new rules and the associated patterns of behavior emerge from the uncoordinated choices of many individuals rather than a single, collective-choice or political process.<sup>8</sup>

<sup>8</sup> Socio-economic evolution differs from biological evolution in that the successful rules, habits, or behavior (generally) spread by imitation and learning, rather than by being passed down directly through replication or sexual reproduction. Socio-economic evolution has also sometimes been described as ‘Lamarckian’, because (unlike in biological evolution), characteristics which were acquired or learned by one individual during their lifetime (rather than inherited) can be passed on to others. Hodgson and Knudsen (2006) argue that this is misleading. They distinguish between the propensity to behave in a particular way (habits and organizational routines), which they view as analogous to a genotype in biology, and the behavior to which the habit or routine gives rise (analogous to biological phenotype). As in biological evolution, it is the success or failure of the behavior (phenotype) that determines whether the habit (genotype) spreads. However, for institutional change to be truly ‘Lamarckian’, they argue,

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 (Williamson, 2000). Depending on the attributes of the particular transaction of interest, some sets of rules (‘governance structures’) will be able to govern this transaction more efficiently than others. TCE *assumes* that the most efficient institutional forms (those which ‘minimize transactions costs’) will emerge – that is, that institutions will develop so as to achieve an optimal ‘match’ with a particular transaction: this is what Williamson refers to as the ‘discriminating alignment’ hypothesis.

Because the outcome is assumed to be determinate (the most efficient institutions will emerge), the process of institutional change is of relatively minor interest in TCE. Implicitly, however, the process of institutional change envisaged is an evolutionary 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. So, for example, if a change in production technology renders existing institutions inefficient, then new, more efficient institutional forms will emerge to replace them.

The TCE 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, while the approach may readily be applied to a choice of contractual forms in an industry with many competitive firms, it becomes more problematic for firms in imperfectly competitive markets; and would seem inappropriate for studying the kinds of deliberately designed rules studied by Libecap and Ostrom, or the development of the US Constitution, for example.

In addition, although Alchian explicitly recognized the potential for the evolutionary process to arrive at a local rather than a global optimum (in other words, the possibility of multiple equilibria), much of the subsequent TCE literature has ignored this possibility. Therefore, the TCE approach has difficulty explaining 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.

Veblen’s (1899) evolutionary theory of institutional change centers on the evolution of ‘habits of thought’, where habits are viewed as durable but (in the

the acquired habits would need to be passed on directly through the inheritance of the genotype, and not indirectly through the behavior itself (the phenotype). The inheritance of acquired or learned characteristics via the phenotype is, however, compatible with Darwinism.

long run) adaptable propensities to think and act in particular ways. The value of habits is a consequence of bounded rationality: they enable the individual to economize on cognitive capacity and interpret information in a complex environment. For Veblen, institutions are ‘in substance, prevalent habits of thought with respect to particular relations and particular functions of the individual and of the community’ (ibid.: 190), and ‘the evolution of social structure has been a process of natural selection of institutions’ – that is, a process of ‘natural selection of the fittest habits of thought’, both through the ‘selection of individuals endowed with the fittest temperament’, and through the ‘adaptation of individual temperament and habits to the changing environment through the formation of new institutions’ (ibid.: 188). Because these habits reside within individuals, institutional change involves the simultaneous co-evolution of both shared prevalent habits of thought (institutions) and the habits of individuals. At any time, therefore, the current habits of thought, both shared and individual, are ‘received from the past’, affected by the present, and together they jointly affect the future path of institutional change.<sup>9</sup>

The notion of habits has been extended to organizations in the form of ‘organizational routines’, which involve ‘the coordination of habits among a group of people such that an orchestrated sequence of actions emerges’ (Knudsen, 2007: 3). In this theory, ‘routines play the role that genes play in biological evolutionary theory’ (Nelson and Winter, 1982: 14), though the process of generating ‘mutations’ is different. Firms may search for new routines if existing routines are unsatisfactory, and they may deliberately attempt to copy successful routines from other firms.<sup>10</sup> However, as the replication of routines involves coordinated changes of several people’s interlocking habits, it will generally require conscious efforts and will rarely be perfect, leading to a constant source of new variation on the original routines (Knudsen, 2007: 14). Similar to Alchian’s theory, evolutionary selection occurs due to the pressure of market competition, as successful firms expand and their routines are copied, whereas unsuccessful firms shrink or fail. However, the difficulty in copying complex routines from one organization to another may slow the diffusion process and create inertia.

Hayek (1973) developed an evolutionary theory of institutional change based on selection at the level of the social group. In this view, rules of conduct ‘have evolved because the groups who practiced them were more successful and displaced others’ (ibid.: 18). As a result, ‘[man’s] thinking and acting are governed by rules which have by a process of selection been evolved in the society in which he lives, and which are thus the product of the experience of generations’ (ibid.: 11). Some of these rules may become formalized, and there may be attempts to design new rules, but Hayek includes these deliberately designed rules and

<sup>9</sup> See Hodgson (2004: chapter 8) for a detailed discussion of Veblen’s theory.

<sup>10</sup> In Nelson and Winter’s theory, these ‘search’ activities are themselves governed by higher-order routines.

organizations, including the state, as forming part of an overall ‘spontaneous order’. According to Hayek, through group selection, this overall configuration of rules will evolve towards an optimal configuration based on consistent general principles (the ‘law of liberty’ based on protection of property rights).

Although he views institutions as rules, Hayek sees shared expectations, rather than rules, as the fundamental source of ‘order’ in society: ‘[the] matching of the intentions and expectations that determine the actions of different individuals is the form in which order manifests itself in social life’ (Hayek, 1973: 36). This brings him close to the ‘Equilibrium View’ of institutions discussed in Section 5 below. Similarly to Veblen, Hayek considers the co-evolution of institutions with the individual ‘mind’, arguing that ‘Mind is as much the product of the social environment in which it has grown up . . . as something that has in turn acted upon and altered these institutions. It is the result of man having developed in society and having acquired those habits and practices that increased the chances of persistence of the group in which he lived’ (ibid.: 17).<sup>11</sup> For Hayek, therefore, the distinction between ‘mind’ and ‘rules’ fades: ‘The mind does not so much make rules as *consists of* rules of action . . . which have come to govern the actions of the individuals because actions in accordance with them have proved more successful than those of competing individuals or groups’ (ibid.: 18, emphasis added). Thus, ‘every man growing up in a given culture will find *in himself* rules . . . (which) are part of a cultural heritage which is likely to be fairly constant’ (ibid.: 19, emphasis added).

The idea of group selection leading to efficient institutions at the societal level seems implausible in many contexts except perhaps over the very long run.<sup>12</sup> And in fact, despite his confidence that evolutionary pressure will eliminate undesirable institutions, Hayek goes on to make a normative argument *against* experimentation, arguing instead that attempts to design institutions should be guided by consistent ideological principles: in particular, that the government should confine itself to protection of property rights, rather than attempt to ‘run the country’. Furthermore, he argues that ‘although rules of just conduct . . . will in the first instance be the product of spontaneous growth, their gradual perfection will require the deliberate efforts of judges . . . who will improve the existing system by laying down new rules’ (ibid.: 100). There is a curious tension

11 Hayek makes clear that the evolution of ‘mind’ happens not through the selection of individuals with innate characteristics, but through the selection of ‘culturally transmitted’ institutions and practices (1973: 23). In contrast, the evolution of human character was envisaged by Veblen as including both the selection of those particular ‘ethnic types’ best suited to current institutions – which seems anachronistic nowadays – and as ‘a process of selective adaptation of habits of thought’ for people of a given ethnic type.

12 Hayek applies it mainly to argue for the superiority of the common law as it developed in England over a period of centuries, as compared with the civil law which developed on the continent. But of course, despite England’s prolonged economic and military success, the allegedly superior common law system never spread to Europe as the group selection argument would seem to imply it eventually ought to.

here between evolution and design: Hayek's contention is that to the extent that intentional human intervention in the design of institutions becomes necessary, it should aim to 'discover' and uphold those ideal rules which (he claims) *would* evolve spontaneously in an evolutionary group-selection model, rather than to invent better rules. Thus, a judge (under common law) becomes 'an institution of a spontaneous order' (ibid.: 95), and rather than a truly evolutionary model (in the Darwinian sense), what emerges is a normative prescription that rules should be designed so as to mimic the outcomes of an evolutionary process.<sup>13</sup>

As we have seen, both TCE and Hayek argue that evolutionary pressure will tend to weed out 'inefficient' institutions, leading to optimal institutional outcomes. However, things become more complicated if we envisage the possibility that the evolutionary process may have multiple equilibria. For example, credit cards are widely used by consumers because many merchants accept them for payment; and they are widely accepted because they are widely used. The resulting equilibrium is associated with a set of formal rules (credit card agreements), norms, and behaviors (carrying little cash), but one might easily envisage alternative institutional configurations, any of which, once established, would generate stable expectations and behavior within the context of a different set of 'rules'; that is, there may be multiple possible sets of self-enforcing rules ('conventions').

In such situations, the question of institutional change revolves around how particular sets of rules are selected. Random chance can certainly play a role. Young (1996) uses an evolutionary framework to argue that historical accidents could lead to the selection of particular conventions, and argues that, in the long run, the pattern of institutional change will follow a 'punctuated equilibrium' process in which rapid switches between conventions are interspersed with long periods of stability.

But the selection of rules need not be entirely random. Sugden (1989) argues that in novel situations, people wishing to coordinate their strategies will generally adopt rules which are analogous to rules with which they are already familiar, such as the 'first come, first served' rule, which can be used to assign property rights in many situations (and thereby avoid potentially costly conflict). Like Hayek, Sugden argues that property rights rules can emerge spontaneously; but unlike Hayek, he does not claim that rules which emerge spontaneously are necessarily efficient (Sugden envisages interactions within a single population, so the 'group selection' argument does not apply).

Knight (1995) also envisages multiple equilibria, but moves beyond thinking of equilibrium selection as purely a coordination problem. He argues that different sets of rules often have different distributional consequences, so different actors

13 There is a strand of Law and Economics literature which claims that the common law will tend to evolve towards efficient rules, because, for example, litigation is more prevalent when the law is inefficient. See Nelson (2005: chapter 3) for a (skeptical) discussion.

may favor the emergence of different rules. In a novel situation, therefore, before the rules which will govern some interaction have become firmly established, people will bargain over which rule to adopt in their individual interactions. This is different from the kind of bargaining envisaged by, for example, Libecap, in that the bargaining takes place at the level of the individual transaction rather than through a centralized political process; but over time, out of this decentralized process, a common convention may emerge. Therefore, if some kinds of actors have greater bargaining power than others,<sup>14</sup> Knight argues that this may systematically affect the kind of rule that ultimately becomes widely used by the society overall.

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. This withdrawal of consent could take the form of organized collective action, but it might also occur through the decentralized actions of many individuals. Bowles and Naidu (2006) develop a model in which players from two populations are randomly matched each period and bargain over the contractual terms which determine the size and distribution of their gains from trade. If a sufficient number of players from one group refuse to accept the status quo contract, this can induce members of the other group to lower their contractual demands, leading to institutional change. As in Young’s model, institutional change will follow a punctuated equilibrium process of periods of stability interspersed with rapid changes.

All of these evolutionary theories, like the design-based theories, regard exogenous parameter changes as a basic source of the impetus for institutional change. In TCE, for example, following a change of parameters, institutions will adapt and a new configuration of optimal institutions will emerge. For Hayek, when circumstances change, judges (through case law) will seek to ‘discover’ how these principles apply to new situations.

In the theories that envision the possibility of multiple evolutionarily stable equilibria however, efficient institutions may not necessarily emerge in response to a parameter change. Instead, institutional change exhibits ‘path-dependence’, in the sense that initial conditions and historical accidents can have a lasting impact on the institutions that are ultimately observed, and it may be possible for inefficient equilibria to arise and persist. For example, an institutional structure which was previously optimal might become sub-optimal as circumstances change, but without a coordinating device, such as legislation or the appearance of a ‘political entrepreneur’, to engineer a change in the rules, the economy might remain stuck in the (now) sub-optimal equilibrium.

<sup>14</sup> There are many possible reasons for differences in bargaining power; Knight focuses on differences in wealth.

In Veblen's theory, changes in population and technology drive institutional change by ensuring that current institutions and habits of thought, inherited from the past, are never ideally suited to the requirements of the present.<sup>15</sup> As a result, institutions and habits are continually evolving, but 'can never catch up with the progressively changing situation in which the community finds itself at any given time'. Thus, 'The evolution of society is substantially a process of mental adaptation on the part of individuals under the stress of circumstances which will no longer tolerate habits of thought formed under and conforming to a different set of circumstances in the past' (Veblen, 1899: 192). With this formulation, in a constantly changing world, there is no guarantee that the institutions observed at any point in time will approximate an 'efficient' configuration, but in Veblen's theory this is because of the speed of adjustment rather than the possibility of getting stuck in a sub-optimal equilibrium.

Ayres (1944) likewise viewed exogenous technological progress as the main driver of institutional change: 'technological development forces change upon the institutional structure by changing the material setting in which it operates' (1944: 187). Ayres, however, viewed institutions as impediments to progress, 'ceremonial' outgrowths of prior technology and a negative, 'past-preserving' influence (*ibid.*: 175). Institutions are resistant to change, in part due to people's emotional attachment to existing institutions, and in part because change threatens existing patterns of status, wealth, and power. For Ayres, therefore, '(t)he history of the human race is that of a perpetual opposition of these forces, the dynamic force of technology continually making for change, and the static force of [institutions] opposing change' (*ibid.*: 176).

Thus, for example, Ayres argues that institutional change did not make the industrial revolution possible. Rather, 'since the industrial revolution did occur, obviously the institutional structure which it confronted was insufficiently solid to prevent change' (1944: 177), and the resulting institutional changes were a consequence rather than a cause of the change in production technologies (*ibid.*: 195). Hodgson (2004) draws a parallel between Ayres' theory and Marxian theories of economic development, in which in which a society's historical development involves a mechanical passage through predictable 'stages' as changes in technology ('modes of production') give rise to changes in institutions ('relations of production'). He also points out that Ayres' and Marx's theories are not 'evolutionary' in the sense of satisfying the key 'Darwinian principles': variation (a source of 'mutations' or experiments), a mechanism of inheritance by which traits are passed on, and a selection process by which successful traits spread.

<sup>15</sup> 'The readjustment of institutions and habitual views to an altered environment is made in response to pressure from without; it is of the nature of a response to stimulus (Veblen, 1899: 192). However, Veblen also allows for the possibility that technological change is itself endogenous, driven in part by current institutions (Brette, 2003: 467–468).



Nelson (2005), like Ayres, views changes in physical technology as the main driver of institutional change, but he allows for feedback from institutions to technology,<sup>16</sup> so that ‘it probably is useful to think of physical and social technologies [i.e. institutions] as coevolving’ (ibid.: 169). However, he argues that in general, ‘all of the conditions associated with sustained and cumulatively significant evolutionary progress tend to be stronger for physical than for social technologies’ (ibid.: 201), in part because of ‘the difficulty of getting sharp, rapid feedback from institutional changes’ (ibid.: 211). Therefore, ‘Social technologies enter the story [of economic growth] largely in terms of how they enable the implementation of physical technologies’ (ibid.: 169), and Nelson worries that institutional change is vulnerable to fads and ideologies which impede efficient institutional change.

#### 4. Blending evolution and design

As we have seen, evolutionary theories tend to neglect the role of collective action and the political process, while theories which view institutional change as the outcome of a centralized collective-choice process have difficulty explaining changes in informal rules (such as social norms) which evolve in a decentralized manner. This is not meant as a criticism of these theories: the theories have been developed to study a variety of different situations, and the assumptions and conclusions naturally reflect these differences. But the question naturally arises as to how to integrate these theories: what is the interaction between spontaneous and deliberate 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 (although there may also be evolutionary processes underlying their creation, as in the case of the common law), and informal rules, which are ‘much more impervious to deliberate policies’ (North, 1990: 6) and therefore (usually) 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

<sup>16</sup> ‘The institutional structure at any time has a profound effect on, and reflects [physical technologies]’ (Nelson, 2005: 163).

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 feedback from lower to higher levels, but then consciously ignores it (2000: 596). Moreover, as defined by Williamson, ‘the New Institutional Economics operates at two levels’ (2000: 610), namely levels 2 and 3, while level 1 institutions are ‘an important but underdeveloped part of the story’ (*ibid.*). 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 collective-choice approach, formal rules change through a political process as a result of deliberate (though boundedly rational) actions by organizations and individual entrepreneurs.<sup>17</sup> The impetus to try to change formal rules can come from exogenous or endogenous parameter changes (including learning).<sup>18</sup> North combines this theory with a theory in which informal rules evolve alongside, and as extensions of, formal rules. Informal rules are reproduced through an evolutionary process of cultural transmission, and 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’ (*ibid.*: 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’ (*ibid.*: 91).

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, organizations develop, and ideologies form in the context of a particular set of formal and informal rules (1990: chapter 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 (*ibid.*: 80–81, 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 that 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), which change slowly because change is continuous, evolutionary,

<sup>17</sup> North (1990: 47) envisages a hierarchy with four levels of formal rules: constitutions, statute and common laws, specific bylaws, and individual contracts.

<sup>18</sup> North identifies two kinds of exogenous parameter shifts: changes in relative prices (including technological change), and changes in preferences.

and decentralized. 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 (i.e. formal rules). Thus, in contrast with North, Roland (2004) makes changes in informal rules, rather than formal rules, the main drivers of institutional change.

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, so that participation in them becomes increasingly mandatory. Thus, Brousseau and Raynaud envision a process by which informal institutions can, in a sense, gradually climb the hierarchical ladder of rules to become more formal and permanent.<sup>19</sup> Bowles and Naidu (2006) similarly highlight the possibility that institutional transitions may begin as decentralized changes in informal practices, which later become confirmed and formalized.

Ruttan (2006) presents a framework in which resource endowments, technology, cultural endowments, and institutions may all mutually affect each other. As a result, disequilibria in resource allocations resulting from changes in resource endowments, technology, or ‘cultural endowments’ can all create a ‘demand’ for institutional change. In some cases, the new institutions may evolve spontaneously; Ruttan uses the example of a Philippine village in which improvements in agricultural technology and population growth created an impetus for changes in land tenure arrangements. In other cases, the new institutions are ‘supplied’ by political entrepreneurs who mobilize political and bureaucratic resources to change the rules. In contrast to Roland’s view of formal rules as ‘fast-changing’, Ruttan argues that the informal land tenancy arrangements in the Philippine village evolved much more quickly through decentralized private contracting between individual farmers than they could have done through deliberate and coordinated institutional change (ibid.: 255). On the other hand, he also notes that other, firmly established and slow-changing informal rules can act as a constraint on changes in formal rules: ‘Cultural endowments, including religion and ideology . . . make some forms of institutional change less costly to establish and impose severe costs on others’ (Ruttan, 2006: 258). This suggests that there may be both fast-moving informal institutions (like the land tenancy arrangements) and slow-moving informal institutions (what Ruttan calls cultural endowments).<sup>20</sup>

<sup>19</sup> The development of Lloyd’s described in Section 2 appears to fit this pattern. Many aspects of commercial law also derived from the codification of merchant’s practices which had evolved spontaneously (Milgrom *et al.*, 1990; Kingston, 2007).

<sup>20</sup> Ruttan states that ‘The distinction that I make between institutions and cultural endowments is that institutions are the *formal* rules and arrangements that govern behavior among and within

## 5. The 'Equilibrium View' of institutions

Despite significantly advancing our understanding of the role of informal rules in institutional change, North still concluded his 1990 book with the thought that 'We need to know much more about [informal rules] and how they interact with formal rules' (North, 1990: 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 Schotter (1981), Calvert (1995), Aoki (2001), Greif and Laitin (2004), Myerson (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.<sup>21</sup> 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. Thus, Calvert (1995: 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

organizations, while cultural endowments are the *informal* codes of behavior that influence individual and group behavior' (2006: 250, emphasis added). However, in the case of the Philippine land tenancy arrangements, he makes clear that the institutional innovation occurred through changes in informal contractual arrangements (p. 255).

21 Hodgson's point (footnote 2) that the state of nature must contain some minimal institutions, such as language, applies here also.

game, and also by the endogenous institutional ‘rules of the game’, which reflect the strategies of the other players (possibly summarized in the form of formal and informal ‘rules’).<sup>22</sup> In maximizing their welfare subject to these constraints, agents choose strategies which, in the aggregate, and perhaps unintentionally, give rise to expectations which 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 expectations (or ‘cultural beliefs’) of many (rational) players about the actions that each will take on and off the path of play (Greif, 2006). Formal rules, however, can also be viewed as devices for coordinating expectations. For example, a formal rule making one player a policeman 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, and how other players will respond: if the rule is effective, then, by virtue of his role, the policeman can take actions and give orders which would not be followed if he were an ordinary citizen. The key point is that the ‘rules’, and the behavior of the enforcer of the rules (policeman), is treated as an endogenous outcome of the analysis. As another example, wars are often governed by ‘rules of war’ (such as an agreement not to use chemical weapons or landmines). These might be formal (the Geneva Conventions), or informal, but, in either case, because of the lack of a higher enforcement authority, they must be self-enforcing to be effective; that is, the behavior specified by the ‘rules’ must be an equilibrium.

In the Equilibrium View, therefore, institutional change becomes fundamentally not about changing rules, but about changing expectations. A new rule which fails to shift people’s expectations in the desired way may have no effect at all (Aoki, 2001: 231); a rule ‘forbidding’ some behavior will be effective only if people generally expect others (including those charged with enforcing the rule) to act in a way which makes it effective. In Ostrom’s terms, it will be a ‘rule-in-form’ but not a ‘rule-in-use’. But more generally, new rules might have an unintended effect. For example, a new law granting the police the power to detain travelers at airports might be expected to enable the police to prevent terrorism, or it might be expected to enable the police to extract bribes by threatening to make people miss their flights. Theories which define institutions as rules tend to obscure these possibilities because they consider the enforcement of rules separately from their content; in the Equilibrium View, in contrast, enforcement is endogenous.

Both deliberate, centralized and evolutionary, decentralized institutional change are compatible with the Equilibrium View. Exogenous parameter shifts

<sup>22</sup> 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).

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 other’s 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 importance of endogenous institutional change. They introduce 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. Thus, institutional change may follow a ‘punctuated equilibrium’ process, in which gradual changes in quasi-parameters occasionally lead to a ‘crisis’ (and institutional change) when it becomes clear that existing patterns of behavior no longer constitute an equilibrium. Similarly, 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.

## **6. Institutional inertia, bounded rationality and the role of history**

As we have seen, various kinds of exogenous or endogenous parameter changes can lead to institutional change. Yet, not all parameter changes lead to institutional change. Whichever definition of institutions we adopt, institutions must be relatively stable and durable if they are to be effective in helping to create order. The durability of institutions makes them meaningful, but it also hinders efforts at institutional change, impedes inter-society transplants of institutions, and makes institutional change ‘overwhelmingly incremental’ (North, 1990: 89). Understanding the sources of institutional stability is therefore crucial to understanding the process of institutional change. Several sources of inertia are identified in the works discussed above.

In the collective-choice framework, free-rider problems can 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: chapter 5) stresses the role of ideology in overcoming the free-rider problem. In some cases, political entrepreneurs may arise to help solve these problems. In Ruttan’s (2006) framework, ‘The supply of major institutional innovations necessarily involves the mobilization of substantial political resources by political entrepreneurs and innovators.’ However, they

will supply these innovations only if the private expected return exceeds the private marginal cost of mobilizing these political resources. As a result, 'To the extent that the private return to political entrepreneurs is different from the social return, the institutional innovation will not be supplied at a socially optimum level' (ibid.: 257).

North (1990) regards informal constraints as the major source of institutional inertia, because they change in an evolutionary manner, and therefore generally slowly. In his view, discrete changes in formal rules often bring about only incremental changes in the overall constraints, because of the persistence of the informal constraints which had evolved as extensions of prior configurations of formal rules. But the influence of informal rules need not be wholly conservative. Ruttan (2006: 258) notes that 'cultural endowments' can actually make some kinds of institutional change easier. He gives the example of Japan, where he argues that 'traditional patterns of cooperation' which emerged in the distant past facilitated modern rural development programs.

Aoki (2001) highlights the existence of strategic linkages and complementarities between institutions in different 'domains' of the economy – for example, between particular kinds of political systems, labor market institutions, and mechanisms of corporate governance. Because institutions in each domain form part of an inter-connected equilibrium whole, radical change is unlikely to occur in a single domain in isolation (and, conversely, institutional change in one domain can have knock-on effects on institutions in other domains).

Institutional inertia is also closely related to the role of history. Most of the theories surveyed above 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. 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.<sup>23</sup> More generally, the resources, physical and human capital, skills, technologies, and organizations accumulated under one set of institutions can gradually alter the set of technologically feasible institutions and thereby affect future institutional development.

Previous institutions also provide focal points which can affect equilibrium selection in novel situations (Sugden, 1989). In the Equilibrium View,

<sup>23</sup> Eggertsson (1996) provides a vivid example: for many centuries, despite great poverty, the inhabitants of Iceland failed to develop a potentially lucrative large-scale fishing industry. The development of the industry was blocked by local elites (landowners) with a vested interest in restricting labor mobility, and by the ruler (the King of Denmark), who feared that the development of an external trade in fish would undermine his political control of the island.

coordinating on one of the many possible institutional equilibria in a society can be viewed as a gigantic multiplayer coordination game, and, as a result, there is no guarantee that efficient institutions will emerge; each actor will rationally follow strategies which constitute a pareto-inferior equilibrium, if s/he 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 (2004) generalize these arguments: knowledge, resource ownership, wealth distribution, and other ‘quasi-parameters’ can all be affected by past institutions, and affect both the future institutional choice set (the set of feasible equilibria) and the choice of institutions within that set.

A potentially important potential source of institutional inertia, emphasized particularly in the ‘old’ institutional economics, is preference endogeneity. As discussed above, Veblen and Hayek saw the evolution of human habits of thought (Veblen) or ‘mind’ (Hayek) as an integral part of institutional change. In Veblen’s theory, institutions (viewed as shared habits of thought) create the social context in which individual habits (dispositions to act in particular ways) are formed. By affecting behavior, institutions affect habits; and ‘by affecting habits, institutions can indirectly influence our purposes or preferences’ (Hodgson, 2004: 187), potentially reinforcing the stability of the institutions themselves as people’s individual preferences and habits are brought into alignment with the ‘shared habits of thought’ which constitute the institution. Thus, according to Hodgson (2007: 331) ‘some malleability of preferences may be necessary to explain fully the evolution and stability of institutions. Institutional stability may be reinforced precisely because of the reconstitutive capacity of institutions to change preferences.’ In this view, institutions shape individuals at the same time as individuals shape institutions.

Bowles (2004: chapter 11) explores the co-evolution of institutions and preferences using evolutionary game theory. In evolutionary game theory models, of course, players do not have ‘preferences’ as such – they have behaviors (or strategies); and behaviors that obtain higher payoffs are more likely to be replicated. However, Bowles interprets the replication process as involving the copying not of behaviors, but of ‘learned norms’ which become internalized as part of an individual’s preferences affecting their choices in particular kinds of situations – a process he refers to as ‘cultural transmission’. People will be more likely to copy and internalize the norms of players whose norms lead them to behave in ways which lead to success.<sup>24</sup> Bowles thereby links changes in preferences with changes in behavior: ‘preferences are endogenous when one’s experiences result in durable changes in one’s behavior in a given situation’ (2004: 378). Thus, the link between behavior and preferences (accomplished

<sup>24</sup> Bowles also allows for strategies to be copied more frequently simply because they are more prevalent in the population (‘conformist transmission’), reflecting social pressures for uniformity.



by Veblen with the notion of habit) is made by Bowles through the notion of internalized norms.

In both the centralized/collective-choice and decentralized/evolutionary theories of institutional change, institutions ultimately arise as a result of purposeful human problem-solving as people 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). For this reason, the nature of bounded rationality and the formation of mental models has become a central concern of recent theorizing on institutional change.

The behavioral economics literature 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 new institutional economics. However, Pagano (2007) distinguishes several different forms of bounded rationality: bounded communications skills, bounded information processing skills, bounded calculation skills, bounded ‘preference formation’ skills (knowing what one wants), and bounded ‘emotional skills’ (ability to control one’s emotions and act rationally).<sup>25</sup> The links between these various forms of bounded rationality at the individual level and institutional change at the aggregate level remain unclear.

In TCE, although bounded rationality is viewed as a key explanation for the existence of institutions, it does not play a major role in institutional change. It matters very little whether players introduce new institutions purposefully, intelligently, or completely at random, because it is assumed that more efficient institutions will drive out less efficient ones through the evolutionary-competitive process. At most, if people do try to design and imitate efficient institutions, this might accelerate the process of institutional adjustment.

In other theories, however, bounded rationality can affect not just the pace but also the direction of institutional change, and can be a source of institutional inertia even in the long run. For example, people may simply be unaware of potentially beneficial institutional changes until the new institutions are ‘invented’. Ruttan (2006) argues that advances in social science knowledge constitute a form of learning which can increase the ‘supply’ of new institutions available to political entrepreneurs. For example, in the 1980s advances in economic theory enabled the design of new rules to create an artificial market for trading in sulphur dioxide emissions (Ruttan, 2006: 259).

<sup>25</sup> Commons, more succinctly, differentiated between ‘passion, stupidity and ignorance’ (Kaufman, 2007: 20).

If some or all of the players attempting to change the rules have an incorrect understanding of the effects of potential changes, then this might affect the political-bargaining process through which institutions are formed in the collective-choice theories (Ostrom, 2005: chapter 4). Risk-aversion exacerbates these problems. Given the uncertainty associated with introducing new institutions, boundedly rational people may be unwilling to experiment, especially with radical institutional changes.

North (2005) presents a framework in which 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. This suggests that a key to understanding institutional change is an understanding of how people learn and revise their ‘mental models’.<sup>26</sup> Therefore, research in cognitive psychology and behavioral economics offers the promise of enabling a deeper understanding of many aspects of institutional change.<sup>27</sup>

North views institutions as rules. Aoki (2007: 6), in contrast, views institutions as ‘self-sustaining, salient patterns of social interactions’ that give rise to ‘common knowledge among the players regarding a particular equilibrium path of the game’ (what we have referred to as the ‘Equilibrium View’), so that institutional change involves a movement to a new equilibrium path.<sup>28</sup> However, because of bounded rationality, each individual agent has an incomplete understanding of this overall equilibrium, and observes only a truncated, simplified version (or mental model) of the overall game as it applies to him or her. 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 mental models 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

26 For example, North (2005: chapter 3) argues that people often reason by analogy.

27 For an optimistic assessment, see Ostrom (2005: chapter 4).

28 Aoki (2007: 6–8) makes clear that these shared beliefs about how the game is to be played may be *represented* by rules, but these rules will be ‘meaningful’ only if they do capture the agents’ equilibrium ‘behavioral beliefs’

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 an organization or political entrepreneur.

## 7. Conclusion

It has not been our goal to discover the ‘best’ theory of institutional change. Perhaps reflecting the differing empirical contexts which the authors of the various theories had in mind, there are substantial differences in emphasis and a variety of ideas about exactly how and how much, for example, history matters, about the role of informal rules and bounded rationality, and about the relative weight of deliberate design and spontaneous evolutionary processes in generating order. This suggests that the appropriate model for studying institutional change, and the answers to the questions we raised in the introduction, may be largely a matter of context. For situations in which competition will tend to weed out inefficient institutions, the transaction cost view is likely to be appropriate. In 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 collective action and political interaction, as in the collective-choice approaches, has proven useful in many real-world settings. However, this approach cannot explain why some formal rules become effective and others do not, and it tends to neglect the role of informal rules. The Equilibrium View of institutions provides a more complete theory by treating both informal and formal rules, and their enforcement, within an integrated framework, and is therefore useful as a broad conceptual framework for understanding institutional change. However, it may introduce unnecessary complexity in the many real-world cases in which formal rules are relatively straightforward and effectively enforced.

A number of issues emerge as priorities for empirical research. One is to develop a sharper understanding of whether and when efficient institutions will tend to emerge. In some situations, the assumption that competition will weed out inefficient institutions may be justifiable, but in others it seems untenable. 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 most of the players better off. Ellickson (1991: chapter 10) argues that even in the case of informal norms which evolve spontaneously, there is a tendency for efficient norms to emerge.

Two other issues stand out. First, the various kinds of ‘informal’ constraints are not always clearly distinguished, leading to considerable potential for confusion. Empirical work is needed to identify how, exactly, the various kinds of

informal constraints work, how they change, how quickly they change, and how they interact with formal rules in both a static and dynamic sense. Second, while many of these theories incorporate some form of bounded rationality, the precise range of roles that various forms of bounded rationality can play in institutional inertia and change remains unclear. Further theoretical and empirical work is needed to clarify the role of bounded rationality, of collective and individual learning, and of endogenous preferences as drivers of, or impediments to, institutional change.

## References

- Acemoglu, D. and J. A. Robinson (2005), 'The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth', *American Economic Review*, 95(3): 546–579.
- Acemoglu, D. and J. A. Robinson (2006), *Economic Origins of Dictatorship and Democracy*, Cambridge: Cambridge University Press.
- Alchian, A. (1950), 'Uncertainty, Evolution and Economic Theory', *Journal of Political Economy*, 58(3): 211–221.
- Alston, L. (1996), 'Empirical Work in Institutional Economics: An Overview', in L. Alston, T. Eggertsson, and D. North (eds.), *Empirical Studies in Institutional Change*, Cambridge University Press, pp. 25–30.
- Aoki, M. (2001), *Towards a Comparative Institutional Analysis*, Cambridge, MA: MIT Press.
- Aoki, M. (2007), 'Endogenizing Institutions and Institutional Changes', *Journal of Institutional Economics*, 3(1): 1–31.
- Ayres, C. E. (1944), *The Theory of Economic Progress*, Chapel Hill, NC: University of North Carolina Press.
- Bowles, S. (2004), *Microeconomics: Behavior, Institutions and Evolution*, Princeton, NJ: Princeton University Press.
- Bowles, S. and S. Naidu (2006), 'Persistent Institutions', working paper, Santa Fe Institute.
- Brette, O. (2003), 'Thorstein Veblen's Theory of Institutional Change: Beyond Technological Determinism', *European Journal of Economic Thought*, 10(3): 455–477.
- Brousseau, E. and E. Raynaud (2008), 'Climbing the Hierarchical, Ladders of Rules: The Dynamics of Institutional Frameworks', working paper, Economix, University of Paris Quest La Défense.
- Caballero, G. (2008), 'El cambio institucional de la economía española del franquismo a la democracia: un análisis histórico institucional', *Política y Gobierno*, V. XV(2): 353–401.
- Calvert, R. (1995), 'Rational Actors, Equilibrium and Social Institutions', in J. Knight and I. Sened (eds.), *Explaining Social Institutions*, Ann Arbor, MI: University of Michigan Press, pp. 57–94.
- Campbell, J. L. (2004), *Institutional Change and Globalization*, Princeton, NJ: Princeton University Press.
- Cooter, R. (1993), 'Against Legal Centrism', *California Law Review*, 81(1): 417–429.
- David, P. A. (1994), 'Why Are Institutions the "Carriers of History"', *Structural Change and Economic Dynamics*, 5(2): 205–220.
- David, P. A. (2007), 'Path Dependence, Its Critics and the Quest for "Historical Economics"', in G. M. Hodgson (ed.), *The Evolution of Economic Institutions: A Critical Reader*, Edward Elgar, pp. 120–142.

- Eggertsson, T. (1996), 'No Experiments, Monumental Disasters: Why It Took a Thousand Years to Develop a Specialized Fishing Industry in Iceland', *Journal of Economic Behavior and Organization*, 30(1): 1–23.
- Ellickson, R. (1991), *Order without Law*, Cambridge, MA: Harvard University Press.
- Greif, A. (2006), *Institutions and the Path to the Modern Economy*, Cambridge: Cambridge University Press.
- Greif, A. and D. Laitin (2004), 'A Theory of Endogenous Institutional Change', *American Political Science Review*, 98(4): 633–652.
- Hayek, F. A. (1973), *Law, Legislation and Liberty, Vol. 1: Rules and Order*, Chicago: University of Chicago Press.
- Hodgson, G. M. (2004), *The Evolution of Institutional Economics: Agency, Structure and Darwinism in American Institutionalism*, Routledge.
- Hodgson, G. M. (2006), 'What Are Institutions?', *Journal of Economic Issues*, 40(1): 1–25.
- Hodgson, G. M. (2007), 'The Revival of Veblenian Institutional Economics', *Journal of Economic Issues*, 16(2): 325–340.
- Hodgson, G. M. and T. Knudsen (2006), 'Dismantling Lamarckism: Why Descriptions of Socio-Economic Evolution as Lamarckian Are Misleading', *Journal of Evolutionary Economics*, 16: 343–366.
- Kandori, M. (1992), 'Social Norms and Community Enforcement', *Review of Economic Studies*, 59: 63–80.
- Kantor, S. E. (1998), *Politics and Property Rights: The Closing of the Open Range in the Postbellum South*, Chicago: University of Chicago Press.
- Kaufman, B. E. (2007), 'The Institutional Economics of John R. Commons: Complement and Substitute for Neoclassical Economic Theory', *Socio-Economic Review*, 5: 3–45.
- Kingston, C. (2007), 'Marine Insurance in Britain and America, 1720–1844: A Comparative Institutional Analysis', *Journal of Economic History*, 67(2): 379–409.
- Knight, J. (1995), 'Models, Interpretations and Theories: Constructing Explanations of Institutional Emergence and Change', in J. Knight and I. Sened (eds.), *Explaining Social Institutions*, Ann Arbor, MI: University of Michigan Press, pp. 95–120.
- Knudsen, T. (2007), 'Organizational Routines in Evolutionary Theory', for *The Handbook of Organizational Routines*, edited by Markus Becker, available at <http://ssrn.com/abstract=997439>.
- Levi, M. (1990), 'A Logic of Institutional Change', in K. Schweers and M. Levi (eds.), *The Limits of Rationality*, Chicago: University of Chicago Press.
- Libecap, G. D. (1989), *Contracting for Property Rights*, Cambridge: Cambridge University Press.
- Milgrom, P., D. North, and B. Weingast (1990), 'The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs', *Economics and Politics*, 1: 1–23.
- Myerson, R. (2004), 'Justice, Institutions, and Multiple Equilibria', *Chicago Journal of International Law*, 5: 91–107.
- Nelson, R. R. (2005), *Technology, Institutions, and Economic Growth*, Cambridge, MA: Harvard University Press.
- Nelson, R. R. and S. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- North, D. (1981), *Structure and Change in Economic History*, New York: W.W. Norton & Co.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

- North, D. (2005), *Understanding the Process of Economic Change*, Princeton, NJ: Princeton University Press.
- Ostrom, E. (2005), *Understanding Institutional Diversity*, Princeton, NJ: Princeton University Press.
- Pagano, U. (2007), 'Bounded Rationality and Institutionalism', in G. M. Hodgson (ed.), *The Evolution of Economic Institutions: A Critical Reader*, Edward Elgar, pp. 19–33.
- Putnam, R. (1993), *Making Democracy Work*, Princeton, NJ: Princeton University Press.
- Roland, G. (2004), 'Understanding Institutional Change: Fast-moving and Slow-moving Institutions', *Studies in Comparative International Development*, 38(4): 109–131.
- Rutherford, M. (1994), *Institutions in Economics: The Old and the New Institutionalism*, Cambridge: Cambridge University Press.
- Ruttan, V. W. (2006), 'Social Science Knowledge and Induced Institutional Innovation: An Institutional Design Perspective', *Journal of Institutional Economics*, 2(3): 249–272.
- Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge: Cambridge University Press.
- Sugden, R. (1989), 'Spontaneous Order', *Journal of Economic Perspectives*, 3(4): 85–97.
- Thelen, K. (2004), *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States, and Japan*, Cambridge: Cambridge University Press.
- Veblen, T. (1899), *The Theory of the Leisure Class: An Economic Study of Institutions*, New York: Macmillan.
- Williamson, O. (2000), 'The New Institutional Economics: Taking Stock, Looking Ahead', *Journal of Economic Literature*, 38: 595–613.
- Wright, C. and C. E. Fayle (1928), *A History of Lloyd's*, London: Macmillan.
- Young, H. P. (1996), 'The Economics of Convention', *Journal of Economic Perspectives*, 10(2): 105–122.